(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 March 2005 (24.03.2005)

PCT

(10) International Publication Number
WO 2005/026362 A2

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, 14/705, 16/18, 16/28, 14/62, G01N 33/68, 33/50, C12Q 1/68, A61K 38/17, 48/00, C12N 15/11

(21) International Application Number:
PCT/US2004/029245

(22) International Filing Date:
8 September 2004 (08.09.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/500,977 8 September 2003 (08.09.2003) US

(71) Applicants *(for all designated States except US)*: **LUDWIG INSTITUTE FOR CANCER RESERACH** [US/US]; 605 tHIRD aVENUE, New York, NY 10158 (US). **LICENTIA LTD.** [FI/FI]; Erottajankatu 19 B, 6th Floor, FIN-00130 Helsinki (FI).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ALITALO, Kari** [FI/FI]; Nyyrikintie 4A, 02100 Espoo (FI). **MAKINEN, Taija** [FI/DE]; Department of Molecule Neurobiology, Max-Planck Institute of Neurobiology, Am Klopferspitz 18A, 82152 Martinsried (DE). **PETROVA, Tatiana** [RU/FI]; Molecular/Cancer Biology Laboratory Biomedicum Helsinki, P.O. Box 63 (Haartmaninkatu 8), FIN-00014 University of Helsinki (FI). **SAHARINEN, Pipsa** [FI/FI]; Molecular/Cancer Biology Laboratory Biomedicum Helsinki, P.O. Box 63 (Haartmaninkatu 8), FIN-00014 Helsinki (FI). **SAHARINEN, Juha** [FI/FI]; Molecular/Cancer Biology Laboratory Biomedicum Helsinki, P.O. Box 63 (Haartmaninkatu 8), FIN-00014 Helsinki (FI).

(74) Agent: **NEVILLE, Katherine, L.**; Marshall, Gerstein & Borun LLP, 233 S. Wacker Drive, 6300 Sears Tower, Chicago, IL 60606-6357 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: LYMPHATIC AND BLOOD ENDOTHELIAL CELL GENES

(57) Abstract: The invention provides polynucleotides and genes that are differentially expressed in lymphatic versus blood vascular endothelial cells. These genes are useful targets for treating diseases involving lymphatic vessels, such as lymphedeman, various inflammatory diseases, and cancer metastasis via the lymphatic system.

WO 2005/026362 A2

LYMPHATIC AND BLOOD ENDOTHELIAL CELL GENES

BACKGROUND OF THE INVENTION

Field of the Invention

The invention relates to polynucleotides and proteins specifically expressed in lymphatic endothelial cells.

Description of the Related Art

Recent evidence on the association of lymphangiogenic growth factors with intralymphatic growth and metastasis of cancers (Mandriota, et al., *EMBO J. 20:*672-682 (2001); Skobe, et al., *Nat. Med. 7*:192-198 (2001); Stacker, et al., *Nat. Med. 7*:186-191 (2001); Karpanen, et al., *Cancer Res. 61*:1786-1790 (2001)) has raised hopes that lymphatic vessels could be used as an additional target for tumor therapy. Cancer cells spread within the body by direct invasion to surrounding tissues, spreading to body cavities, invasion into the blood vascular system (hematogenous metastasis), as well as spread via the lymphatic system (lymphatic metastasis). Regional lymph node dissemination is the first step in the metastasis of several common cancers and correlates highly with the prognosis of the disease. The lymph nodes that are involved in draining tissue fluid from the tumor area are called sentinel nodes, and diagnostic measures are in place to find these nodes and to remove them in cases of suspected metastasis. However, in spite of its clinical relevance, little is known about the mechanisms leading to metastasis via the bloodstream or via the lymphatics.

Until recently, the lymphatic vessels have received much less attention than blood vessels, despite their importance in medicine. Lymphatic vessels collect protein-rich fluid and white blood cells from the interstitial space of most tissues and transport them as a whitish opaque fluid, the lymph, into the blood circulation. Small lymphatic vessels coalesce into larger vessels, which drain the lymph through the thoracic duct into large veins in the neck region. Lymph nodes serve as filtering stations along the lymphatic vessels and lymph movement is propelled by the contraction of smooth muscles surrounding collecting lymphatic vessels and by bodily

movements, the direction of flow being secured by valves as it is in veins. The lymphatic capillaries are lined by endothelial cells, which have distinct junctions with frequent large interendothelial gaps. The lymphatic capillaries also lack a continuous basement membrane, and are devoid of pericytes. Anchoring filaments connect the abluminal surfaces of lymphatic endothelial cells to the perivascular extracellular matrix and pull to maintain vessel patency in the presence of tissue edema. The absence or obstruction of lymphatic vessels, which is usually the result of an infection, surgery, or radiotherapy and in rare cases, a genetic defect, causes accumulation of a protein-rich fluid in tissues, lymphedema. The lymphatic system is also critical in fat absorption from the gut and in immune responses. Bacteria, viruses, and other foreign materials are taken up by the lymphatic vessels and transported to the lymph nodes, where the foreign material is presented to immune cells and where dendritic cells traverse via the lymphatics. There has been slow progress in the understanding of and ability to manipulate the lymphatic vessels.

Abnormal development or function of the lymphatic ECs can result in tumors or malformations of the lymphatic vessels, such as lymphangiomas or lymphangiectasis. Witte, et al., *Regulation of Angiogenesis* (eds. Goldber, I.D. & Rosen, E.M.) 65-112 (Birkäuser, Basel, Switzerland, 1997). The VEGFR-3 tyrosine kinase receptor is expressed in the normal lymphatic endothelium and is upregulated in many types of vascular tumors, including Kaposi's sarcomas. Jussila, et al., *Cancer Res 58*, 1955-1604 (1998); Partanen, et al., *Cancer 86*:2406-2412 (1999). Absence or dysfunction of lymphatic vessels which can result from an infection, surgery, radiotherapy or from a genetic defect, causes lymphedema, which is characterized by a chronic accumulation of protein-rich fluid in the tissues that leads to swelling. The importance of VEGFR-3 signaling for lymphangiogenesis was revealed in the genetics of familial lymphedema, a disease characterized by a hypoplasia of cutaneous lymphatic vessels, which leads to a disfiguring and disabling swelling of the extremities. Witte, et al., *Regulation of Angiogenesis* (eds. Goldber, I.D. & Rosen, E.M.) 65-112 (Birkäuser, Basel, Switzerland, 1997); Rockson, S.G., *Am. J. Med. 110*, 288-295 (2001). Some members of families with lymphedema are heterozygous for missense mutations of the *VEGFR3* exons encoding the tyrosine

kinase domain, which results in an inactive receptor protein. Karkkainen, et al., *Nature Genet. 25*:153-159 (2000); Irrthum, et al., *Am. J. Hum. Genet. 67*:295-301 (2000).

There is a need in the art for information on the transcriptional program which controls the diversity of endothelial cells, and into the mechanisms of angiogenesis and lymphangiogenesis. There is also a need in the art for new vascular markers, which may be used as valuable targets in the study of a number of diseases involving the lymphatic vessels, including tumor metastasis.

SUMMARY OF THE INVENTION

The compositions of the present invention include isolated polynucleotides and polypeptides, in particular, lymphatic endothelial genes, isolated polypeptides encoded by these polynucleotides, recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, and antibodies that specifically recognize one or more epitopes present on such polypeptides.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

In selected embodiments, such isolated polynucleotides of the invention represent a polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a fragment thereof that encodes a corresponding polypeptide.

The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes to the complement of the nucleotide sequence of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 under highly stringent hybridization conditions; a polynucleotide that hybridizes to the complement of the nucleotide sequence of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 under moderately stringent

hybridization conditions; a polynucleotide which is an allelic variant of any polynucleotide recited above; a polynucleotide which encodes a species homologue of any of the proteins recited above; or a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptide encoded by any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41. Exemplary high stringency hybridization conditions are hybridization at 42°C for 20 hours in a solution containing 50% formamide, 5xSSPE, 5x Denhardt's solution, 0.1% SDS and 0.1 mg/ml denatured salmon sperm DNA, with a wash in 1xSSC, 0.1% SDS for 30 minutes at 65°C.

Another aspect of the invention is drawn to LEC and BEC polypeptides, including polypeptides encoded by the polynucleotides described above. In some embodiments, the polypeptides are the mature forms of the polypeptides of the invention. Expressly contemplated is a purified and isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46; and a purified and isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43; and (b) an extracellular domain fragment of at least 10 amino acids of an amino acid sequence of (a). Further, this aspect of the invention includes a purified and isolated, soluble polypeptide as described immediately above, comprising an extracellular domain fragment of an amino acid sequence selected from the group consisting of : SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43 wherein the polypeptide lacks any transmembrane domain. Such a polypeptide may further lack any intracellular domain. Also, the invention contemplates a fusion protein comprising a polypeptide as described above fused, e.g., to an immunoglobulin fragment comprising an immunoglobulin constant region.

Another aspect of the invention is an isolated polypeptide comprising an amino acid sequence at least 95% identical to amino acids 1-1005 of SEQ ID NO: 43. In a related aspect, a composition is contemplated which comprises a soluble polypeptide comprising a fragment of one of the amino acid sequences disclosed herein that comprises a functional protein domain. One exemplary embodiment

comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 43, wherein said fragment lacks the transmembrane and intracellular amino acids from approximately residues 1006-1028 of SEQ ID NO: 43. Further contemplated is an isolated polypeptide comprising an adhesion domain having an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 20 or SEQ ID NO: 43, wherein said fragment includes at least one vonWillebrand factor domain or at least one Cache domain.

In an additional aspect, the invention contemplates polypeptide fragments of any one of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46, wherein the fragment comprises at least one protein domain of the given polypeptide sequence, as set out herein. The invention further contemplates polypeptide fragments comprising an amino acid sequence at least 95% identical to a fragment of any one of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46 comprising a protein domain of the polypeptide set forth herein. Exemplary embodiments include, e.g., polypeptide fragments comprising a protein domain that is a protein kinase domain selected from the group consisting of a protein kinase domain set out in SEQ ID NO: 8, a protein kinase domain set out in SEQ ID NO: 26, a protein kinase domain at approximately amino acids 88-342 of SEQ ID NO: 42, and a protein kinase domain at approximately amino acids 51-317 of SEQ ID NO: 46; a Zinc binding domain selected from the group consisting of a Zinc-binding domain set out in SEQ ID NO: 36, a zinc-binding domain at approximately amino acids 627-561 of SEQ ID NO: 6 and a zinc-binding domain at approximately amino acids 1030-1094 of SEQ ID NO: 45.

Another aspect of the invention is drawn to an isolated polypeptide at least 95% identical to a polypeptide set forth herein, comprising a particular biological activity associated with a protein domain as set forth herein. Exemplary biological activities include ubitquitin ligase activity of a Hect domain in SEQ ID NO: 28; guanine-nucleotide-dissociation-stimulating activity of at least one RCC1-like domain in SEQ ID NO: 28; cation transport activity of a cation transport domain set forth in SEQ ID NO: 12; GTP-binding activity of at least one GTP-binding

domain set forth in SEQ ID NO: 32, SRPY domain activity of at least one SPRY domain of SEQ ID NO: 6, and Kelch domain activity of at least one Kelch domain of SEQ ID NO: 30 or SEQ ID NO: 43. In another aspect, the invention contemplates an isolated polypeptide comprising an amino acid sequence at least 95% identical to a transcription factor selected from the group consisting of SEQ ID NO: 30, SEQ ID NO: 38, SEQ ID NO: 40, and SEQ ID NO: 44, or a protein involved in cytoskeletal regulation selected from the group consisting of SEQ ID NO: 18, SEQ ID NO: 36 and SEQ ID NO: 45.

In a related aspect, the invention provides a composition comprising a polynucleotide or polypeptide or protein as described above and a pharmaceutically acceptable diluent, carrier or adjuvant. Polypeptide compositions of the invention may comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier. Further provided is a kit comprising such a composition and a protocol for administering the pharmaceutical composition to a mammalian subject to modulate the lymphatic system in the subject. The invention also provides an antibody that specifically binds to a polypeptide as described above, and that antibody is humanized in some embodiments. Antibodies contemplated by the invention include, for example, monoclonal and polyclonal antibodies, single chain antibodies, chimeric antibodies, bifunctional/bispecific antibodies, humanized antibodies, human antibodies, single and multiple domain antibodies and complementarity-determining region (CDR)-grafted antibodies, including compounds which include CDR sequences that specifically recognize a polypeptide of the invention, provided that such antibodies each specifically bind at least one polypeptide according to the invention. Still further, the invention provides a protein comprising an antigen binding domain of an antibody that specifically binds to a polypeptide as described hereinabove, wherein the protein specifically binds to the polypeptide.

The invention also relates to methods for producing a polypeptide comprising growing a culture of the cells of the invention in a suitable culture medium, and purifying the protein from the culture or from an extract of the cells. In particular, the invention contemplates a method for producing a LEC polypeptide comprising steps of growing a host cell transformed or transfected with an expression

vector as described herein under conditions in which the cell expresses the polypeptide encoded by the polynucleotide.

Methods of identifying the products and compositions described herein are also provided by the invention. In particular, the invention provides a method of identifying a LEC nucleic acid comprising: (a) contacting a biological sample containing a candidate LEC nucleic acid with a polynucleotide comprising a fragment of at least 14 contiguous nucleotides of a sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a complement thereof, under the following stringent hybridization conditions: (i) hybridization at 42°C for 20 hours in a solution containing 50% formamide, 5xSSPE, 5x Denhardt's solution, 0.1% SDS and 0.1 mg/ml denatured salmon sperm DNA, and (ii) washing for 30 minutes at 65°C in 1xSSC, 0.1% SDS; and (b) detecting hybridization of the candidate LEC nucleic acid and the polynucleotide, thereby identifying a LEC nucleic acid.

The invention also provides a method of identifying a LEC protein comprising: (a) contacting a biological sample containing a candidate LEC protein with a LEC protein binding partner selected from the group consisting of an antibody as described herein or a protein or polypeptide as described herein, under conditions suitable for binding therebetween; and (b) detecting binding between the candidate LEC protein and the LEC binding partner, thereby identifying a LEC protein.

Another related aspect of the invention is a method of identifying a LEC comprising: (a) contacting a biological sample comprising cells with a LEC binding partner under conditions suitable for binding therebetween, wherein the LEC binding partner comprises an antibody that binds to a polypeptide comprising a sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43, or comprises an antigen binding fragment of the antibody; and (b) identifying a LEC by detecting binding between a cell and the LEC binding partner, where binding of the LEC binding partner to the cell identifies a LEC.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as primers for PCR, use for

chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, such as a lymphatic endothelial cell, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g.*, *in situ* hybridization.

In another aspect, the invention provides a composition comprising an isolated polynucleotide that comprises a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46 and a pharmaceutically acceptable diluent, carrier or adjuvant. In some embodiments, the composition comprises a polynucleotide that comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a fragment thereof that encodes the polypeptide.

Still another aspect of the invention is an expression vector comprising an expression control sequence operably linked to a polynucleotide that comprises a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46. In some embodiments, the expression vector is a replication-deficient adenoviral or adeno-associated viral vector containing the polynucleotide. A related aspect of the invention is a composition comprising an expression vector as described above and a pharmaceutically acceptable diluent, carrier, or adjuvant. Further, the invention provides a kit comprising the composition containing either the above-described polynucleotide or vector and a pharmaceutically acceptable diluent, carrier or adjuvant, packaged with a protocol for administering the composition to a mammalian subject to modulate the lymphatic system in the subject.

The invention further provides a host cell transformed or transfected with an expression vector as described above.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. In addition, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide.

In one aspect of the invention, a method is provided for differentially modulating the growth or differentiation of blood endothelial cells (BEC) or lymphatic endothelial cells (LEC), comprising contacting endothelial cells with a composition comprising an agent that differentially modulates blood or lymphatic endothelial cells, said agent selected from the group consisting of: (a) a polypeptide that comprises an amino acid sequence of a BEC polypeptide or a LEC polypeptide, or an active fragment of the polypeptide ; (b) a polynucleotide that comprises a nucleotide sequence that encodes a polypeptide according to (a); (c) an antibody that specifically binds to a polypeptide according to (a); (d) a polypeptide comprising a fragment of the antibody of (c), wherein the fragment and the antibody bind to the polypeptide; (e) an antisense nucleic acid to a human gene or mRNA encoding the polypeptide of (a); (f) an interfering RNA (RNAi) to a human gene or mRNA encoding the polypeptide of (a). In a related aspect, the method of the invention provides an antibody, a peptide or a small molecular weight compound for differentially modulating the growth or differentiation of blood endothelial cells (BEC) or lymphatic endothelial cells (LEC). The method may involve endothelial cell contact with the composition *ex vivo* or *in vivo*. The composition may comprise a pharmaceutically acceptable diluent, adjuvant, or carrier, and the contacting step may comprise administering the composition to a mammalian subject to differentially modulate BECs or LECs in the mammalian subject.

Further, the method may comprise identifying a human subject with a disorder characterized by hyperproliferation of LECs; and administering to the human subject the composition, wherein the agent differentially inhibits LEC growth compared to BEC growth; alternatively the method may comprise identifying a human subject with a disorder characterized by hyperproliferation of LECs; screening LECs of the subject to identify overexpression of a polypeptide set forth in any of Tables 1 and 2 herein; and administering to the human subject the composition,

wherein the agent differentially inhibits LEC growth compared to BEC growth by inhibiting expression of the polypeptide identified by the screening step.

This aspect of the invention also contemplates a method of modulating the growth of lymphatic endothelial cells in a human subject, comprising steps of identifying a human subject with a hypoproliferative lymphatic disorder; screening the subject to identify underexpression or underactivity of a LEC polypeptide set forth in any of Table 1 or Table 2 herein; administering to the human subject the composition, wherein the agent comprises the LEC polypeptide (a) identified by the screening step or an active fragment of the polypeptide, or comprises the polynucleotide (b) that comprises a nucleotide sequence that encodes the polypeptide.

A related aspect of the invention is drawn to a use of an agent for the manufacture of a medicament for the differential modulation of blood vessel endothelial cell (BEC) or lymphatic vessel endothelial cell (LEC) growth or differentiation, the agent selected from the group consisting of: (a) a polypeptide that comprises an amino acid sequence of a BEC polypeptide or a LEC polypeptide, or an active fragment of the polypeptide ; (b) a polynucleotide that comprises a nucleotide sequence that encodes a polypeptide according to (a); (c) an antibody that specifically binds to a polypeptide according to (a); (d) a polypeptide comprising a fragment of the antibody of (c), wherein the fragment and the antibody bind to the polypeptide; (e) an antisense nucleic acid to a human gene or mRNA encoding the polypeptide of (a); (f) an interfering RNA (RNAi) to a human gene or mRNA encoding the polypeptide of (a). It is further contemplated that an antibody, a peptide or a small molecular weight compound may be useful for the manufacture of a medicament for the differential modulation of blood endothelial cell (BEC) or lymphatic endothelial cell (LEC) growth or differentiation. In some embodiments, the medicament may promote growth and/or differentiation of a BEC or LEC; in other embodiments, the medicament inhibits growth and/or differentiation of such a cell.

In another aspect, the invention provides a method of identifying compounds that modulate growth of endothelial cells, comprising culturing endothelial cells in the presence and absence of a compound; and measuring expression of at least one BEC or LEC gene in the cells, wherein the gene is a gene

set forth in any of Tables 1 and 2 herein, wherein a change in expression of at least one BEC gene in the presence compared to the absence of the compound identifies the compound as a modulator of BEC growth, and wherein a change in expression of at least one LEC gene in the presence compared to the absence of the compound identifies the compound as a modulator of LEC growth. The method may be used to screen for a compound that selectively modulates BEC or LEC growth or differentiation, wherein the measuring step comprises measuring expression of at least one BEC gene and at least one LEC gene in the cells, and wherein the method comprises screening for a compound that selectively modulates BEC or LEC growth or differentiation by selecting a compound that differentially modulates expression of the at least one BEC gene compared to expression of the at least one LEC gene. In a related aspect, the method further comprises a second BEC or LEC gene that encodes a polypeptide set forth in any of Tables 3 and 4 of PCT/US03/06900.

Further, the invention comprehends a method or use according to the aspects of the invention described above, wherein the polypeptide is a LEC polypeptide selected from a gene found in any of Tables 1 and 2 herein, and the agent differentially modulates LEC growth or differentiation over BEC growth or differentiation. In some embodiments, the LEC polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43; in other embodiments, the LEC polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 20 and SEQ ID NO: 32. In these embodiments, an agent may be an antibody that specifically binds to a LEC polypeptide as described above, or a polypeptide fragment of such an antibody. Further, the agent may be an extracellular domain of a polypeptide described above, a polynucleotide encoding an extracellular domain, or an antisense molecule or nucleic acid. Alternatively, the polypeptide is a BEC polypeptide selected from the BEC polypeptides set forth in any of Tables 1 and 2 herein, and the agent differentially modulates BEC growth or differentiation over LEC growth or differentiation. It is further contemplated that the above methods further comprise screening LECs of the subject to identify overexpression of a second polypeptide set forth in any of Tables 3 and 4 of PCT/US03/06900 (specifically incorporated herein by reference).

The methods of the present invention further relate to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of a prognostic and/or diagnostic evaluation of disorders as recited above and for the identification of subjects exhibiting a predisposition to such conditions. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders related to lymphatic endothelial cells.

The invention also provides methods for the identification of compounds that modulate the expression of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders related to expression of proteins encoded by any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 as recited above. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention.

Further, the invention provides a method of assaying for risk of developing hereditary lymphedema, comprising (a) assaying nucleic acid of a human subject for a mutation that correlates with a hereditary lymphedema phenotype and alters the encoded amino acid sequence of at least one gene allele of the human subject when compared to the amino acid sequence of the polypeptide encoded by a corresponding wild-type gene allele, wherein the wild-type polypeptide is a polypeptide identified in any of Tables 1 and 2 herein. Alternatively, a method of assaying for risk of developing hereditary lymphedema, comprises (a) assaying nucleic acid of a human subject for a mutation that correlates with a hereditary lymphedema phenotype and alters the encoded amino acid sequence of at least one gene allele of the human subject when compared to the amino acid sequence of the polypeptide encoded by a corresponding wild-type gene allele, wherein the wild-type polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43; (b) correlating the presence or absence of the mutation in the nucleic acid to a risk of developing hereditary

lymphedema, wherein the presence of the mutation in the nucleic acid correlates with an increased risk of developing hereditary lymphedema, and wherein the absence of the mutation in the nucleic acid correlates with no increased risk of developing hereditary lymphedema.

In another method of assaying for risk of developing hereditary lymphedema, the steps comprise (a) assaying nucleic acid of a human subject for a mutation that alters the encoded amino acid sequence of at least one intracellular protein or transcription factor allele of the human subject and alters transcription modulation activity of the transcription factor polypeptide encoded by the allele, when compared to the transcription modulation activity of a transcription factor polypeptide encoded by a wild-type allele, wherein the wild-type intracellular protein or transcription factor polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 18, 26, 28, 30, 32, 34, 36, 38, 40, 44, 45, and 46; and (b) correlating the presence or absence of the mutation in the nucleic acid to a risk of developing hereditary lymphedema, wherein the presence of the mutation in the nucleic acid correlates with an increased risk of developing hereditary lymphedema, and wherein the absence of the mutation in the nucleic acid correlates with no increased risk of developing hereditary lymphedema. As one example, the assaying identifies a mutation altering activity of a transactivating or DNA binding domain of a Sox18 allele. In some embodiments of the method, the mutation reduces transcriptional activation of a SOX18-responsive gene compared to transcriptional activation of the gene in the absence of said mutation.

In a related aspect, the invention provides a method of assaying for risk of developing hereditary lymphedema, comprising (a) assaying nucleic acid of a human subject for a mutation that alters the encoded amino acid sequence of at least one LEC gene allele of the human subject and alters the binding affinity of the adhesion polypeptide encoded by the LEC gene allele, when compared to the binding affinity of an adhesion polypeptide encoded by a wild-type allele, wherein the wild-type adhesion polypeptide comprises an amino acid sequence of SEQ ID NO: 20; and (b) correlating the presence or absence of the mutation in the nucleic acid to a risk of developing hereditary lymphedema, wherein the presence of the mutation in the

nucleic acid correlates with an increased risk of developing hereditary lymphedema, and wherein the absence of the mutation in the nucleic acid correlates with no increased risk of developing hereditary lymphedema.

In the methods of assaying for risk of developing hereditary lymphedema according to the invention, the assaying may identify the presence of the mutation, and the correlating step may identify the increased risk of the patient developing hereditary lymphedema.

A related method according to the invention is a method of screening a human subject for an increased risk of developing hereditary lymphedema comprising assaying nucleic acid of a human subject for a mutation that alters the encoded amino acid sequence of at least one polypeptide comprising an amino acid sequence found in any of Tables 1 and 2. In some embodiments of this method, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46 in a manner that correlates with the risk of developing hereditary lymphedema, and it is expressly contemplated that the polypeptide may comprise the amino acid sequences set forth in SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43.

A related aspect of the invention is drawn to methods of assaying or screening for risk of developing hereditary lymphedema as described above, wherein the method comprises at least one procedure selected from the group consisting of: (a) determining a nucleotide sequence of at least one codon of at least one polynucleotide of the human subject; (b) performing a hybridization assay to determine whether nucleic acid from the human subject has a nucleotide sequence identical to or different from one or more reference sequences; (c) performing a polynucleotide migration assay to determine whether nucleic acid from the human subject has a nucleotide sequence identical to or different from one or more reference sequences; and (d) performing a restriction endonuclease digestion to determine whether nucleic acid from the human subject has a nucleotide sequence identical to or different from one or more reference sequences.

A related aspect of the invention provides methods of assaying or screening for risk of developing hereditary lymphedema as described above, wherein

the method comprises: performing a polymerase chain reaction (PCR) to amplify nucleic acid comprising the coding sequence of the LEC polynucleotide, and determining nucleotide sequence of the amplified nucleic acid.

Further provided by the invention is a method of screening for a hereditary lymphedema genotype in a human subject, comprising: (a) providing a biological sample comprising nucleic acid from said subject, and (b) analyzing the nucleic acid for the presence of a mutation altering the encoded amino acid sequence of the at least one allele of at least one gene in the human subject relative to a human gene encoding an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46, wherein the presence of a mutation altering the encoded amino acid sequence in the human subject in a manner that correlates with lymphedema in human subjects identifies a hereditary lymphedema genotype. In some embodiments of this method, the biological sample is a cell sample. In other embodiments of this method, the analyzing comprises sequencing a portion of the nucleic acid. In still further embodiments of this method, the human subject has a hereditary lymphedema genotype identified by the method of screening.

In is contemplated that the above methods further comprise assaying the nucleic acid for a second mutation that correlates with a hereditary lymphedema phenotype and alters the encoded amino acid sequence of a second gene allele of the human subject when compared to the amino acid sequence of the polypeptide encoded by a corresponding wild-type allele of said second gene, wherein the wild-type polypeptide encoded by said second gene is a polypeptide identified in any of Tables 3 and 4 of PCT/US03/06900.

Another aspect of the invention provides a method of inhibiting lymphangiogenesis comprising administering to a subject an inhibitor of a LEC transmembrane polypeptide, wherein the LEC transmembrane polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43, and wherein the inhibitor is selected from the group consisting of (a) a soluble extracellular domain fragment of the LEC transmembrane polypeptide; (b) an antibody that binds to the extracellular domain of the LEC

transmembrane polypeptide; (c) a polypeptide comprising an antigen binding domain of the antibody according to (b); and (d) an antisense nucleic acid complementary to the nucleic acid encoding the LEC transmembrane polypeptide or its complement. In some embodiments of the method, the inhibitor is a polypeptide comprising an extracellular domain fragment of an LEC polypeptide. It is further contemplated that the sequence of said extracellular domain is selected from the group consisting of amino acids 1-1005 of SEQ ID NO: 43. In some embodiments of the method, the subject is a human containing a tumor.

In a related aspect, the invention provides a method for modulating lymphangiogenesis in a mammalian subject comprising: administering to a mammalian subject in need of modulation of lymphangiogenesis an antisense molecule to a LEC polynucleotide, in an amount effective to inhibit transcription or translation of the polypeptide encoded by the LEC polynucleotide, wherein the LEC polynucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41. Antisense molecules of suitable sequence and length are known in the art, and include sequences perfectly complementary to a coding region of a coding strand of a nucleic acid, as well as lengths capable of relatively specific and stable hybridization to a target *in vivo* (e.g., lengths of each incremental integer in the range of 8-50, with a preferred length of at least 17 nucleotides).

The methods of the invention also include methods for the treatment of disorders related to lymphatic endothelial cells as recited above which may involve the administration of such compounds to individuals exhibiting symptoms or tendencies related to such disorders.

In another aspect, the invention provides a method of treating hereditary lymphedema, comprising: (a) identifying a human subject with hereditary lymphedema and with a mutation that alters the encoded amino acid sequence of at least one polypeptide of the human subject, relative to the amino acid sequence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46; and (b) administering to the subject a lymphatic growth factor

selected from the group consisting of a VEGF-C polynucleotide, a VEGF-C polypeptide, a VEGF-D polynucleotide, and a VEGF-D polypeptide.

The invention also provides a method of treating hereditary lymphedema comprising: identifying a human subject with lymphedema and with a mutation in at least one allele of a gene encoding a LEC protein identified in any of Tables 1 and 2, wherein the mutation correlates with lymphedema in human subjects, and with the proviso that the LEC-specific protein is not VEGFR-3; and administering to the subject a composition comprising a lymphatic growth agent selected from the group consisting of VEGF-C polypeptides, VEGF-D polypeptides, VEGF-C polynucleotides, and VEGF-D polynucleotides. The invention also comprehends use of a lymphatic growth agent selected from the group consisting of VEGF-C polypeptides, VEGF-D polypeptides, VEGF-C polynucleotides, and VEGF-D polynucleotides in the manufacture of a medicament for the treatment of hereditary lymphedema resulting from a mutation in a LEC gene identified in any of Tables 1 and 2. It is contemplated that the method further comprises a mutation in at least one allele of a second gene encoding a LEC protein identified in any of Tables 3 and 4 of PCT/US03/06900, wherein the mutation in said second gene correlates with lymphedema in human subjects.

In addition, the invention encompasses methods for treating such diseases or disorders by administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can effect such modulation either at the level of target gene expression or target protein activity. These treatment methods include the administration of a polypeptide or a polynucleotide according to the invention to an endothelial cell, e.g., a LEC and/or a BEC, or to an organism such as a human patient. An exemplary method according to this aspect of the invention is the administration of a therapeutic selected from the group consisting of an antisense polynucleotide capable of modulating the expression of at least one polynucleotide according to the invention, a polypeptide according to the invention, a polynucleotide according to the invention, an antibody or antibody fragment specifically recognizing a polypeptide according to the invention, a VEGF-

C polynucleotide, a VEGF-C polypeptide, a VEGF-D polynucleotide, a VEGF-D polypeptide and a soluble VEGFR-3 polypeptide.

In another aspect, the invention provides a method of screening for an endothelial cell disorder or predisposition to the disorder, comprising obtaining a biological sample containing endothelial cell mRNA from a human subject; and measuring expression of a BEC or LEC gene from the amount of mRNA in the sample transcribed from the gene, wherein the BEC or LEC gene encodes a polypeptide identified in any of Tables 1 and 2. In a related aspect, the method further comprises a second BEC or LEC gene that encodes a polypeptide set forth in any of Tables 3 and 4 of PCT/US03/06900.

The invention also relates to a method of inhibiting the growth of a lymphatic endothelial cell, the method comprising contacting the cell with a composition comprising at least one antibody conjugated to an agent capable of inhibiting the growth, wherein the agent is selected from the group consisting of a cytotoxic agent and a cytostatic agent, and wherein the antibody specifically binds to a polypeptide encoded by a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 19, 21 and 23. In specific embodiments of this method, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43.

The invention further relates to methods of detecting a lymphatic endothelial cell, the method comprising contacting the cell with a composition comprising at least one antibody conjugated to a detectable agent, such as a fluorescent molecule or a radiolabeled molecule. In specific embodiments, the antibody specifically binds to a polypeptide encoded by a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 19, 21 and 23. In further specific embodiments of this method, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43.

The invention still further relates to methods of isolating a lymphatic endothelial cell, comprising contacting the cell with a solid matrix comprising at least

one antibody capable of binding to a transmembrane protein in the cell membrane of the cell, and isolating cells specifically bound to the antibody matrix. In specific embodiments, the antibody specifically binds to a polypeptide encoded by a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 19, 21 and 23. In further specific embodiments of this method, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43.

The invention also relates to the administration of an agonist or antagonist to a lymphatic endothelial cell surface receptor, comprising selecting an antibody, a peptide or a small molecular weight compound that is capable of specifically binding to a lymphatic endothelial cell-specific protein, wherein the antibody, peptide or small molecular weight compound is an agonist or antagonist for a growth factor receptor, a cytokine receptor, a chemokine receptor, or a hemopoietic receptor, and contacting the antibody, peptide or small molecular weight compound with the lymphatic endothelial cell in need of growth stimulation or inhibition. In specific embodiments, such lymphatic endothelial cells are associated with lymphedema, lymphangioma, lymphangiomyeloma, lymphangiomatosis, lymphangiectasis, lymphosarcoma, and lymphangiosclerosis.

The invention also relates to the administration of a cytotoxic or cytostatic drug to a cancerous lymphatic endothelial cell, comprising selecting an antibody, a peptide or a small molecular weight compound that is capable of specifically binding to a lymphatic endothelial cell-specific protein, wherein the antibody, peptide or small molecular weight compound is complexed to the cytotoxic or cytostatic drug. In specific embodiments, administration of such complexes is useful in the treatment of malignant tumor diseases prone to metastatic spread through the lymphatic system.

The invention also provides a method of monitoring the efficacy or toxicity of a drug on endothelial cells, comprising the steps of measuring expression of at least one BEC or LEC gene in endothelial cells of a mammalian subject before and after administering a drug to the subject, wherein the at least one LEC or BEC gene encodes a polypeptide found in any of Tables 1 and 2 herein, and wherein

changes in expression of the LEC or BEC gene correlates with efficacy or toxicity of the drug on endothelial cells. It is contemplated that the method further comprises a second BEC or LEC gene that encodes a polypeptide set forth in any of Tables 3 and 4 of PCT/US03/06900.

The invention further relates to a method of detecting a lymphatic endothelial cell, comprising contacting said cell with an antibody as described herein, wherein the antibody is detectably labeled.

The invention still further relates to a method of inhibiting at least one biological activity of a lymphatic endothelial cell, comprising contacting the cell with an agent capable of binding to at least one polypeptide encoded by any one of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43, wherein the activity of the polypeptide is reduced relative to the activity of a polypeptide that is not contacted with the agent.

The invention also relates to a method of inhibiting the growth of a lymphatic endothelial cell, the method comprising contacting the cell with an antisense oligonucleotide capable of specifically binding to at least one polynucleotide selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41. In a specific embodiment, the antisense oligonucleotide consists essentially of about 12 to about 25 contiguous nucleotides of any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

It is further contemplated that the methods of the invention, such as identifying a LEC nucleic acid; modulating the growth of lymphatic endothelial cells in a human subject; identifying compounds that modulate growth of endothelial cells; identifying a human subject with a disorder characterized by hyperproliferation of LECs; assaying for risk of developing hereditary lymphedema; screening for an increased risk of developing hereditary lymphedema; screening for a hereditary lymphedema genotype in a human subject; and inhibiting lymphangiogenesis, may be carried out using compositions comprising polypeptide fragments as described above. In some embodiments, the polypeptide is at least 95% identical to a polypeptide set forth herein, comprising a particular biological activity associated with a protein

domain as set forth herein. The polypeptide may have transcription factor activity or cytoskeletal regulation activity. Additional exemplary biological activities include the ubitquitin ligase activity of a Hect domain, guanine-nucleotide-dissociation-stimulating activity of at least one RCC1-like domain, cation transport activity of a cation transport domain, GTP-binding activity of at least one GTP-binding domain, SPRY domain activity of at least one SPRY domain, and Kelch domain activity of at least one Kelch domain. Variations on the methods above may also be carried out using the above-described polypeptide fragment compositions.

Additional features and variations of the invention will be apparent to those skilled in the art from the entirety of this application, and all such features are intended as aspects of the invention.

DETAILED DESCRIPTION OF THE INVENTION

A major role of the lymphatic vasculature is to remove an excess of the protein-rich interstitial fluid that continuously escapes from the blood capillaries, and to return it to the blood circulation (Witte, M.H., et al., *Microsc. Res. Tech. 55*:122-145. 2001; Karpanen, T., et al., *J. Exp. Med. 194*:F37-F42. 2001; Karkkainen, M.J., *et al., Trends Mol. Med.* 7:18-22, 2001). In addition, the lymphatic system provides constant immune surveillance by filtering lymph and its antigens through the chain of lymph nodes, and also serves as one of the major routes for absorption of lipids from the gut. It has been known for a long time that in many types of cancer the lymphatic vessels provide a major pathway for tumor metastasis, and regional lymph node dissemination correlates with the progression of the disease. Hereditary lymphedema, post-surgical secondary lymphedema and lymphatic obstruction in filariasis, are all characterized by disabling and disfiguring swelling of the affected areas, linked to the insufficiency or obstruction of the lymphatics. Witte, M.J., et al., *Microsc. Res. Tech 55*:122-145 (2001).

In spite of the importance of lymphatic vessels in medicine, the cell biology of this part of the vascular system has received little attention until recently. Studies during the past four years have uncovered the existence of the lymphatic

specific vascular endothelial growth factors VEGF-C and VEGF-D, which serve as ligands for the receptor tyrosine kinase VEGFR-3, and demonstrated their importance for the normal development of the lymphatic vessels (*See*, Jeltsch *et al., Science 276*:1423-1425 (1997); Veikkola, T., *et al., EMBO J. 20*:1223-1231 (2001); Mäkinen, T., *et al., Nat. Med.* 7:199-205 (2001)). These molecules also appear to be involved in the development of lymphedema and lymphatic metastasis (Karpanen *et al., J. Exp. Med. 194*:F37-F42 (2001); Karkkainen et al., *Trends Mol. Med.* 7:18-22, 2001).

The growth factor Vascular Endothelial Growth Factor C (VEGF-C), as well as native human, non-human mammalian, and avian polynucleotide sequences encoding VEGF-C, and VEGF-C variants and analogs, have been described in detail in International Patent Application Number PCT/US98/01973, filed February 2, 1998 and published on August 6, 1998 as International Publication Number WO 98/33917; in Joukov *et al., J. Biol. Chem.*, 273(12): 6599-6602 (1998); and in Joukov *et al., EMBO J.*, 16(13): 3898-3911 (1997), all of which are incorporated herein by reference in their entirety. As explained therein in detail, human VEGF-C (SEQ ID NOS: 1 and 2) is initially produced in human cells as a prepro-VEGF-C polypeptide of 419 amino acids. A cDNA encoding human VEGF-C (SEQ ID NO: 1) has been deposited with the American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110-2209 (USA), pursuant to the provisions of the Budapest Treaty (Deposit date of 24 July 1995 and ATCC Accession Number 97231). VEGF-C sequences from other species also have been reported. See Genbank Accession Nos. MMU73620 (*Mus musculus*); and CCY15837 (*Coturnix coturnix*) for example, incorporated herein by reference.

The prepro-VEGF-C polypeptide is processed in multiple stages to produce a mature and most active VEGF-C polypeptide of about 21-23 kD, as assessed by SDS-PAGE under reducing conditions (SEQ ID NO: 2). Such processing includes cleavage of a signal peptide (residues 1-31); cleavage of a carboxyl-terminal peptide (corresponding approximately to amino acids 228-419 and having a pattern of spaced Cysteine residues reminiscent of a Balbiani ring 3 protein (BR3P) sequence [Dignam *et al., Gene*, 88:133-40 (1990); Paulsson *et al., J. Mol. Biol.*, 211:331-49 (1990)]) to produce a partially-processed form of about 29 kD; and cleavage

(apparently extracellularly) of an amino-terminal peptide (corresponding approximately to amino acids 32-103) to produced a fully-processed mature form of about 21-23 kD. Experimental evidence demonstrates that partially-processed forms of VEGF-C (e.g., the 29 kD form) are able to bind VEGFR-3 (Flt4 receptor), whereas high affinity binding to VEGFR-2 occurs only with the fully processed forms of VEGF-C. It appears that VEGF-C polypeptides naturally associate as non-disulfide linked dimers.

It has been demonstrated that amino acids 103-227 of VEGF-C are not all critical for maintaining VEGF-C functions. A polypeptide consisting of amino acids 113-213 (and lacking residues 103-112 and 214-227) retains the ability to bind and stimulate VEGF-C receptors, and it is expected that a polypeptide spanning from about residue 131 to about residue 211 will retain VEGF-C biological activity. The Cysteine residue at position 156 has been shown to be important for VEGFR-2 binding ability. However, VEGF-C ΔC_{156} polypeptides (i.e., analogs that lack this Cysteine due to deletion or substitution) remain potent activators of VEGFR-3. The Cysteine at position 165 of VEGF-C polypeptide is essential for binding either receptor, whereas analogs lacking the Cysteine at positions 83 or 137 compete with native VEGF-C for binding with both receptors and stimulate both receptors.

VEGF-D is structurally and functionally most closely related to VEGF-C (*see* U.S. Patent 6,235,713 and International Patent Publ. No. WO 98/07832, incorporated herein by reference). See SEQ ID NO: 3 for the polynucleotide sequence of VEGF-D; the encoded amino acid sequence is set forth in SEQ ID NO: 4. Like VEGF-C, VEGF-D is initially expressed as a prepro-peptide that undergoes N-terminal and C-terminal proteolytic processing, and forms non-covalently linked dimers. VEGF-D stimulates mitogenic responses in endothelial cells *in vitro*. During embryogenesis, VEGF-D is expressed in a complex temporal and spatial pattern, and its expression persists in the heart, lung, and skeletal muscles in adults. Isolation of a biologically active fragment of VEGF-D designated VEGF-DΔNΔC, is described in International Patent Publication No. WO 98/07832, incorporated herein by reference. VEGF-DΔNΔC consists of amino acid residues 93 to 201 of VEGF-D (SEQ ID NO: 4) optionally linked to the affinity tag peptide FLAG®, or other sequences.

The prepro-VEGF-D polypeptide has a putative signal peptide of 21 amino acids and is apparently proteolytically processed in a manner analogous to the processing of prepro-VEGF-C. A "recombinantly matured" VEGF-D lacking residues 1-92 and 202-354 retains the ability to activate receptors VEGFR-2 and VEGFR-3, and appears to associate as non-covalently linked dimers. Thus, preferred VEGF-D polynucleotides include those polynucleotides that comprise a nucleotide sequence encoding amino acids 93-201. The guidance provided above for introducing function-preserving modifications into VEGF-C polypeptides is also suitable for introducing function-preserving modifications into VEGF-D polypeptides. As another aspect of the invention, practice of the invention methods is contemplated wherein VEGF-D polypeptides are employed in lieu of VEGF-C polypeptides.

When compared with the blood vascular endothelium, the lymphatic endothelium exhibits specific morphological and molecular characteristics. For example, the lymphatic capillaries are larger than blood capillaries, they have an irregular or collapsed lumen with no red blood cells, a discontinuous basal lamina, overlapping intercellular junctional complexes and anchoring filaments that connect the lymphatic endothelial cells to the extracellular matrix (Witte, M.H., et al., *Microsc. Res. Tech. 55*:122-145 (2001)). Unlike the blood capillaries, the lymphatic capillaries lack pericyte coverage. At the molecular level several lymphatic specific markers have been identified, including VEGFR-3, the Prox-1 transcription factor, the hyaluronan receptor LYVE-1, the membrane mucoprotein podoplanin, the beta-chemokine receptor D6, the cytoskeletal proteins desmoplakin I and II and the macrophage mannose receptor I (Wigle, J.T. & Oliver, G., *Cell 98*:769-778 (1999); Banerji, S., et al., *J. Cell Biol. 144*:789-801 (1999): Breiteneder-Geleff, S., et al., *Am. J. Pathol. 154*:385-394 (1999): Nibbs, R.J., et al., *Am. J. Pathol. 158*:867-877 (2001); Ebata, N., et al., *Microvasc. Res. 61*:40-48. (2001); Irjala, H., et al., *J. Exp. Med. 194*:1033-1041 (2001)). The present invention relates to the genetic identity of lymphatic capillary endothelial cells versus blood vascular endothelial cells using a gene profiling approach.

"Stringent hybridization conditions" or "stringent conditions" refer to conditions under which a nucleic acid such as an oligonucleotide will specifically

hybridize to its target sequence. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer nucleic acids hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration conditions) at which 50% of the nucleic acids complementary to the target sequence hybridize to the target sequence at equilibrium. The term "complementary" refers to standard Watson-Crick base pairing between nucleotides of two nucleic acid molecules. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and at a temperature that is at least about 30°C for short probes, primers or oligonucleotides (*e.g.,* 10 to 50 nucleotides) and at least about 60°C for longer probes, primers or oligonucleotides. Stringent conditions also can be achieved with the addition of destabilizing agents, such as formamide, as is known in the art Exemplary stringent hybridization conditions are hybridization at 42°C for 20 hours in a solution containing 50% formamide, 5xSSPE, 5x Denhardt's solution, 0.1% SDS and 0.1 mg/ml denatured salmon sperm DNA, with a wash in 1xSSC, 0.1% SDS for 30 minutes at 65°C.

Distinct gene expression profiles for blood vascular and lymphatic endothelial cells have been discovered (See PCT/US03/06900, incorporated herein by reference). These results provide new insights into the phenotypic diversity of endothelial cells and reveal new potential lymphatic endothelial cell molecules, some of which could provide important targets for the therapy of diseases characterized by abnormal angiogenesis or lymphangiogenesis.

Differences in the expression of genes encoding proteins involved in inflammatory processes were found, as well as in those mediating cell-cell and cell-matrix interactions. Furthermore, several previously unknown genes were identified in the context of endothelial cell biology, which were differentially expressed in the two cell lineages. Several of the genes were originally cloned from neural tissues, including genes involved in the uptake of synaptic macromolecules and in synapse

formation and remodeling (neuronal pentraxins I and II (Kirkpatrick, L.L., et al., *J. Biol. Chem. 275*:17786-17792. 2000), in the trafficking of synaptic vesicles (NAP-22 (Yamamoto, Y., et al., *Neurosci. Lett. 224*:127-130. 1997), piccolo (Fenster, S.D., et al., *Neuron 25*:203-214 (2000)) and in the axon growth and guidance (Nr-CAM (Grumet, M., *Cell Tissue Res. 290*:423-428 (1997), reelin (Rice, D.S. & Curran, T., *Annu. Rev. Neurosci. 24*:1005-1039 (2001)).

In addition, the LECs especially expressed a number of uncharacterized genes, which were originally cloned and highly expressed in nervous tissues (KIAA genes (Kikuno, R., et al., *Nucleic Acids Res. 30*:166-168. 2002)). The gene expression profiling data disclosed herein therefore support the view that the same molecular mechanisms that are involved in governing neural cell positioning, in guiding axonal growth cones to their specific targets and in synaptogenesis may also be commonly used in the development of the vascular system and in the establishment of BEC and LEC identity. Some other signaling molecules first described in the developing nervous system have already been implicated in the development of the vasculature and vice versa (Shima and Mailhos, *Curr. Opin Genet. Dev. 10*:536-542 (2000); Oosthuyse, et al., *Nat. Genet. 28*:131-138 (2001); Sondell, et al., *Eur. J. Neurosci. 12*:4243-4254 (2000)).

PCT/US03/06900 identified expression of several genes in the LECs, previously shown to be expressed in smooth muscle cells (SMCs) and pericytes, such as matrix Gla, a mineral binding extracellular matrix protein involved in the inhibition of vascular and tissue calcification (Luo, G., et al., *Nature 386*:78-81 (1997)), monoamine oxidase A, the main degradative enzyme of monoamine hormones and neurotransmitters (Rodriguez, M.J., et al., *Cell Tissue Res. 304*:215-220 (2001)), integrin $\alpha 9$ (Palmer, E.L., et al., *J. Cell Biol. 123*:1289-1297 (1993)) and apolipoprotein D (Hu, C.Y., et al., *J. Neurocytol. 30*:209-218 (2001)). Some similarity of gene expression patterns between LECs and SMCs could be related to the lack of SMC around lymphatic capillaries. Instead, LECs may carry out some SMC functions by themselves. For example, lymph flow is maintained due to the intrinsic contractility of the LECs (Witte, M.H., et al., *Microsc. Res. Tech. 55*:122-145 (2001)), reminiscent of the ability of vascular SMCs to contract.

Molecular discrimination of the lymphatic and blood vessels is essential in studies of diseases involving the blood and/or lymphatic vessels and in the targeted treatment of such diseases. To date, several lymphatic endothelial specific markers have been identified, but some of them are expressed only in a subset of the lymphatic vessels, while others also occur in some blood vessel endothelia or in other cell types and their expression patterns may change in pathological conditions (for example, VEGFR-3 (Valtola, R., et al., *Am. J. Pathol. 154*:1381-1390. 1999)). Identification of new vascular markers according to the invention should provide a more reliable analysis of the blood and lymphatic vessels in pathological situations and eventually better diagnosis and treatment. Furthermore, inhibition of the function of certain molecules involved in the regulation of angiogenesis and/or lymphangiogenesis is known to prevent tumor growth and metastasis, and stimulation of the growth of blood or lymphatic vessels has been shown to be beneficial in several pathological conditions. Thus the LEC specific molecular regulators identified according to the invention may provide new targets for the treatment of diseases characterized by abnormal angiogenesis and lymphangiogenesis.

Several of the new LEC genes encode transmembrane proteins which are expected to be specific molecular markers for lymphatic endothelial cells (Table 1). These genes and encoded proteins are useful for targeted treatment of diseases that involve lymphatic vessels. They may also be useful for preparing antibodies, as antibodies against LEC-specific proteins can be used to discriminate between blood and lymphatic vessels in pathological and physiological situations. Antibodies may also be useful for the isolation of lymphatic endothelial cells. These proteins may also play a role in the regulation of lymphangiogenesis, and can provide new candidate genes for diseases that involve lymphatic vessels, such as lymphedema.

The lymphatic endothelial cell specific surface molecules can be used for molecular drug targeting with antibodies, peptides and small-molecular weight compounds, which can act as agonists or antagonists for growth factor receptor, cyto- and chemokine receptor, and hemopoietin receptor signaling, cell adhesion and cell interaction with extracellular matrix or with other cell surface molecules. Such molecules can also be used for targeting of cytotoxic or cytostatic drugs into the

lymphatic endothelial cells and for the attachment of electron-dense, radio-opaque or radioactive markers for imaging of disease processes associated with the lymphatic vessels. Such diseases include lymphedema, lymphangioma, lymphangiomyoma, lymphangiomatosis, lymphangiectasis, lymphosarcoma and lymphangiosclerosis.

The lymphatic endothelial cell surface molecules may be used for targeting of gene therapy for example by antibody-coated liposomes (containing proteins or genes as cargo) or by viral transducing vectors such as adenoviruses, adeno-associated viruses or lentiviruses having modified capsid/envelope proteins. The manipulation of lymphatic endothelial cell specific molecules may be applicable to treatment of disease processes associated with tissue edema by increasing fluid transport across the lymphatic vessel wall for example by modulating endothelial cell-cell or cell-matrix interactions or via stimulating transendothelial transport. Targeting of the lymphatic endothelial cells for example with cytotoxic or cytostatic compounds is contemplated to be valuable in malignant tumor diseases prone to metastatic spread via the lymphatic system.

The lymphatic endothelial cell molecules may allow the improved *in vitro* growth of lymphatic endothelial cells as well as *in vitro* tissue engineering of lymphatic vessels for use in diseases where the lymphatics have been damaged, such as after surgery and in various forms of lymphedema. Ligands of the cell surface proteins may further be applied to coat various polymeric matrices for the adhesion of cells in, e.g., bioimplants.

The lymphatic endothelial-cell-specific molecules such as surface molecules can provide important tools for the modulation of inflammatory, autoimmune and infectious processes involving leukocyte migration and immune recognition as well as the stimulation of secondary immune responses. Such processes include the migration of antigen presenting cells into the lymphatic system including lymph nodes as well as transendothelial cell trafficking of lymphocytes and other leukocyte subclasses and the homing, survival and function of the various classes of leukocytes.

These molecules may allow one to modulate the metabolism of fatty acids including fatty acid/chylomicron absorption from the gut and regulation of fat

accumulation in adipose tissue in various organs such as in the subcutaneous tissue and in the arterial wall.

Lymphatic endothelial-cell-specific molecules may further allow one to modulate the metabolism of fatty acids including fatty acid/chylomicron absorption from the gut and regulation of fat accumulation in adipose tissue in various organs such as in the skin subcutaneous tissue and in the arterial wall.

The lymphatic-cell-specific transmembrane proteins are expected to function in cell adhesion (e.g., adhesion between lymphatic endothelial cell-lymphatic endothelial cell, lymphatic endothelial cell-smooth muscle cell, lymphatic endothelial cell-immune system cell such as lymphocyte or dendritic cell), cell-extracellular matrix contacts, or as receptors such as growth factor, cytokine, chemokine or microbial receptors or ion channels. The transmembrane (TM) proteins connect to intracellular molecules that can induce cell growth, cell migration, cell apoptosis, cell differentiation or cell adhesion or other cellular functions specific for endothelial cells such as expression of adhesion receptors for leukocytes, release of nitric oxide, antigoagulant proteins, uptake of fluid and proteins from surrounding tissues and fat from gut or adipose tissues. TM proteins with short intracellular domains can function as auxiliary receptors in complex with other TM proteins.

The transmembrane proteins and their intracellular binding partner molecules can be used as molecular markers for lymphatic endothelial cells in normal and disease conditions, and to discriminate between blood and lymphatic vessels in pathological and physiological situations.

Antibodies against lymphatic specific transmembrane proteins, as well as peptides and small molecular compounds binding to extracellular domains of lymphatic-specific TM proteins can be used for the attachment of electron-dense, radio-opaque or radioactive markers for imaging of disease processes associated with the lymphatic vessels. Such diseases include lymphedema, lymphangioma, lymphangiomyoma, lymphangiomatosis, lymphangiectasis, lymphosarcoma and lymphangiosclerosis. Similarly, the lymphatic vessels can be visualized, e.g., during therapy of patients suffering from insufficient lymphatic growth, such as in lymphedema, or alternatively during treatment aiming to prevent lymphatic growth,

e.g., in tumors, thereby facilitating the monitoring of the therapeutic method of the invention.

Antibodies against LEC-specific TM proteins are also expected to be useful for the isolation of lymphatic endothelial cells.

Antibodies against lymphatic-specific transmembrane proteins, or peptides or small-molecule compounds binding to the extracellular domain of lymphatic-specific TM proteins are expected to be useful in targeting drug delivery to lymphatic endothelial cells, e.g., by coupling an antibody, peptide or small-molecule compound to a cytotoxic or cytostatic compound. Such coupled compounds are useful as therapeutics in the treatment of malignant tumor diseases prone to metastatic spread via the lymphatic system, as well as in ameliorating a symptom associated with any such disease. The antibodies, peptides or small-molecule compounds can also be coupled to stimulatory lymphatic endothelial molecules such as growth factors, cytokines and chemokines to promote stimulation.

Additionally, antibodies against lymphatic-specific TM proteins or peptides, or small-molecule compounds binding to the extracellular domain of lymphatic-specific TM proteins, may be used for targeting of gene therapy, for example, by antibody-coated liposomes (containing proteins, genes or other molecules as cargo) or by viral transducing vectors such as adenoviruses, adeno-associated viruses, lentiviruses, or the like, having modified capsid/envelope proteins The manipulation of lymphatic endothelial-cell-specific molecules are expected to be applicable to the treatment of disease processes associated with tissue edema due to the relative absence, or relative dysfunction, of lymphatic vessels, which can result from an infection, surgery, radiotherapy or a genetic defect by increasing fluid transport across the lymphatic vessel wall, for example by modulating endothelial cell-cell or cell-matrix interactions or by stimulating transendothelial transport.

The lymphatic endothelial cell molecules are expected to improve the *in vitro* growth of lymphatic endothelial cells, as well as the *in vitro* tissue engineering of lymphatic vessels for use in treating disorders or diseases where the lymphatics have been damaged, such as after surgery, in various forms of lymphedema, and in other applications as described herein. Ligands of the cell-surface proteins may

further be applied as a coating to various polymeric matrices for the adhesion of cells in, e.g., bioimplants.

Inflammatory, autoimmune and infectious processes involving leukocyte migration and immune recognition, such as migration of antigen-presenting cells into the lymphatic system, including lymph nodes, as well as transendothelial cell trafficking of lymphocytes and other leukocyte subclasses and the homing, survival and function of the various classes of leukocytes can be modulated by targeting endothelial-cell-specific TM proteins, which mediate these cell adhesion processes.

Upregulation of lymphatic-specific genes in, *e.g.*, cancer are expected to be useful as diagnostic markers, and monitoring such upregulated expression with an antibody against a lymphatic endothelial-cell-specific protein, *e.g.*, by immunostaining of tissue(s) or by using a probe hybridizable to a lymphatic endothelial-cell-specific mRNA, *e.g.*, under stringent hybridization conditions as described herein, is contemplated.

Lymphatic endothelial-cell-specific transcription factors are expected to be useful for the differentiation of lymphatic endothelial cells from embryonic stem cells, endothelial precursor cells, or blood vascular endothelial cells.

The lymphatic endothelial transcription factors are expected to improve the *in vitro* growth of lymphatic endothelial cells, as well as to facilitate *in vitro* tissue engineering of lymphatic vessels for use in treating disorders or diseases where the lymphatics have been damaged, such as after surgery, in various forms of lymphedema, and in other applications disclosed herein.

Intracellular signaling proteins participating in signaling pathways regulating lymphatic endothelial cell proliferation, differentiation, apoptosis, migration or adhesion are expected to be useful targets for small-molecule compounds inhibiting these signaling events, and cellular functions dependent on such signaling. Signaling proteins are also expected to participate in VEGFR-3 signaling pathways, and will be useful in modulating cellular activities controlled, at least in part, by VEGFR-3 signaling, such as lymphangiogenesis.

The lymphatic endothelial cell molecules are expected to improve the *in vitro* growth of lymphatic endothelial cells as well as *in vitro* tissue engineering of lymphatic vessels for use in treating diseases or disorders where the lymphatics have been damaged, such as after surgery, in various forms of lymphedema, and in other applications as described herein.

Lymphatic-specific transcription factors are also expected to be useful in modulating gene expression in endothelial cells to induce the expression of other lymphatic-specific genes in, for example, blood vascular endothelial cells or endothelial precursor cells.

Lymphatic-specific gene transcripts are expected to provide useful targets for RNA interference (RNAi)-induced inhibition of expression. RNAi technology is expected to be useful in the methods according to the invention, such as therapeutic methods effective in treating hyper- and hypo-proliferative endothelial-cell-associated diseases and disorders, as well as methods of ameliorating a symptom of any such disease or disorder. RNAi methodologies are known in the art and known RNAi technologies are contemplated as useful in various aspects of the invention. *See* Fire et al., *Nature* 391:806-811. (1998) and Sharp, P., *Genes and Dev.* 13:139-141. (1999), each of which is incorporated herein by reference. It is preferred that RNAi compounds be double-stranded RNA molecules corresponding to part or all of a coding region of a desired target for expression.

As noted, several of the new LEC genes encode transcription factors, may control cellular fate and have an important role in the differentiation of lymphatic endothelial cells. Transcription factors disclosed herein (Table 2) may control transcription of genes involved for example in the proliferation of lymphatic endothelial cells, and may be important molecular regulators of lymphatic growth. Lymphatic endothelial cell specific transcription factors can be used for the differentiation of lymphatic endothelial cells from embryonic stem cells, endothelial precursor cells or from blood vascular endothelial cells.

The lymphatic endothelial transcription factors may allow the improved *in vitro* growth of lymphatic endothelial cells as well as in vitro tissue

engineering of lymphatic vessels for use in diseases where the lymphatics have been damaged, such as after surgery and in various forms of lymphedema.

Polynucleotides of the Invention

In general, the isolated polynucleotides of the invention include the LEC and BEC polynucleotides exhibiting differential expression and identified in Tables 1 and 2, including sequences identifiers of these polynucleotides and their known database accession numbers, where applicable. The polynucleotide sequences may include a coding region and may include non-coding flanking sequences, which are readily identifiable by one of skill in the art. The invention contemplates polynucleotides comprising part, or all, of a coding region, with or without flanking regions, e.g., poly A sequences, 5' non-coding sequences, and the like. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes to the complement of the nucleotide sequence of any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 under highly stringent hybridization conditions; a polynucleotide that hybridizes to the complement of the nucleotide sequence of any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 under moderately stringent hybridization conditions; a polynucleotide which is an allelic variant of any polynucleotide recited above; a polynucleotide which encodes a species homologue of any of the proteins recited above; or a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptide of any one of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46. Such polynucleotides hybridize under the above conditions to the complement of any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 or to a fragment of any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 wherein the fragment is greater than at least about 10 bp, and, in alternate embodiments, is about 20 to about 50 bp, or is greater than about 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, or 800 bp, where appropriate. References to "bp" refer to the length of double-, or

single-, stranded nucleic acids in nucleotides or base pairs, as would be apparent from the context.

The polynucleotides of the invention also provide polynucleotides that are variants of the polynucleotides recited above. Typically, such a variant sequence varies from one of those listed herein by no more than about 20%, *i.e.*, the number of individual nucleotide substitutions, additions, and/or deletions in a similar sequence, as compared to the corresponding reference sequence, divided by the total number of nucleotides in the variant sequence is about 0.2 or less. Such a sequence is said to have 80% sequence identity to the listed sequence. Such a variant sequence can be routinely identified by applying the foregoing algorithm.

In one embodiment, a variant polynucleotide sequence of the invention varies from a listed sequence by no more than 10%, *i.e.*, the number of individual nucleotide substitutions, additions, and/or deletions in a variant sequence, as compared to the corresponding reference sequence, divided by the total number of nucleotides in the variant sequence is about 0.1 or less. Such a sequence is said to have 90% sequence identity to the listed sequence. Such a variant sequence can be routinely identified by applying the foregoing algorithm.

In an alternate embodiment a variant sequence of the invention varies from a listed sequence by no more than by no more than 5%, *i.e.*, the number of individual nucleotide substitutions, additions, and/or deletions in a variant sequence, as compared to the corresponding reference sequence, divided by the total number of nucleotides in the variant sequence is about 0.05 or less. Such a sequence is said to have 95% sequence identity to the listed sequence. Such a variant sequence can be routinely identified by applying the foregoing algorithm.

In yet another alternate embodiment, a variant sequence of the invention varies from a listed sequence by no more than 2%, *i.e.*, the number of individual nucleotide substitutions, additions, and/or deletions in a variant sequence, as compared to the corresponding reference sequence, divided by the total number of nucleotides in the variant sequence is about 0.02 or less. Such a sequence is said to have 98% sequence identity to the listed sequence. Such a variant sequence can be routinely identified.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook, J. et al. (2d Ed.; 1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). Useful nucleotide sequences for joining to polypeptides include an assortment of vectors, *e.g.*, plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, sequencing vectors, and retroviral vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic variations thereof. Allelic variations can be routinely determined by comparing the polynucleotide sequence provided in any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, a representative intermediate fragment thereof, or a nucleotide sequence at least 99.9% identical to any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific open reading frames (ORFs) disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Unless provided for otherwise here, all terms are defined as is known in the art, for example as employed in U.S. Patent No. 6,350,447, incorporated herein by reference.

Also contemplated are antisense polynucleotides based on the sequence of any of the LEC or BEC polynucleotides set forth found in any of Tables 1 and 2 herein or Tables 3 and 4 of PCT/US03/06900. Such antisense polynucleotides are substantially complementary (e.g., at least 90% complementarity), and preferably perfectly complementary, to any one of the polynucleotides of the invention, or fragments thereof, set out in any of Tables 1 and 2 herein or Tables 3 and 4 of PCT/US03/06900 (incorporated herein by reference) that are differentially expressed in LECs and BECs. These polynucleotide sequences include any of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a fragment thereof comprising at least 10 contiguous nucleotides. Antisense nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). Methods for designing and optimizing antisense nucleotides are described in Lima *et al.*, (*J Biol Chem,* ;272:626-38. 1997) and Kurreck *et al.*, (*Nucleic Acids Res.*, ;30:1911-8. 2002). In one aspect, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "conceding region" of the coding strand of a nucleotide sequence encoding the polynucleotide. The term "conceding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of the mRNA of the

polynucleotide of the invention, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of the mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize or bind to cellular mRNA and/or genomic DNA encoding the complementary polynucleotide, thereby inhibiting expression of the protein (e.g., by inhibiting transcription and/or translation). The hybridization can reflect conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix.

An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). Additional routes of antisense therapy may be used in the invention, e.g., topical administration, transdermal administration [reviewed by Brand in *Curr. Opin. Mol. Ther.* 3:244-8, 2001] antisense administration using nanoparticulate systems [Lambert et al., *Adv. Drug. Deliv. Rev.* 47:99-112, 2001], or administration of antisense

nucleotides conjugated with peptide [Juliano et al., *Curr. Opin. Mol. Ther.* 2:297-303, 2000].

The invention further contemplates use of the polynucleotides of the invention for gene therapy or in recombinant expression vectors which produce polynucleotides or polypeptides of the invention that can regulate an activity of LEC genes, and are useful in therapy of LEC disorders such as lymphedema. Delivery of a functional gene encoding a polypeptide of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, including viral vectors (*e.g.*, adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, *Nature*, supplement to vol. 392, no. 6679, pp. 25-20 (1998). For additional reviews of gene therapy technology see Friedmann, (*Science,* 244: 1275-1281. 1989); Verma, (*Scientific American*: 263:68-72, 81-84. 1990); and Miller, (*Nature,* 357: 455-460. 1992). Introduction of any one of the nucleotides of the present invention or a gene encoding a polypeptide of the invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on, or activity in, such cells. In another embodiment, cells comprising vectors expressing the polynucleotides or polypeptides of the invention may be cultured *ex vivo* and administered to an individual in need of treatment for an LEC disease or disorder.

Given the foregoing disclosure of the nucleic acid constructs, it is possible to produce the gene product of any of the genes comprising the sequence of any of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 by routine recombinant DNA/RNA techniques. A variety of expression vector/host systems may be utilized to contain and express the coding sequence. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, phagemid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (*e.g.*, baculovirus); plant cell systems transfected with virus expression vectors (*e.g.*, Cauliflower Mosaic Virus, CaMV; Tobacco Mosaic

Virus, TMV) or transformed with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmid); or even animal cell systems. Mammalian cells that are useful in recombinant protein productions include, but are not limited to, VERO cells, HeLa cells, Chinese hamster ovary (CHO) cells, COS cells (such as COS-7), WI38, BHK, HepG2, 3T3, RIN, MDCK, A549, PC12, K562 and HEK 293 cells.

Polypeptides of the invention

In general, the isolated LEC polypeptides of the invention are encoded by the above-described differentially expressed LEC polynucleotides of the invention. The sequences of the LEC polypeptides are provided in Table 1 and 2, associated with their known database accession numbers, where applicable. The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NOS.: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46 or an amino acid sequence encoded by any one of the nucleotide sequences set forth in SEQ ID NOS.: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46, or the corresponding full length or mature protein. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NOS.: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46, or the corresponding full length or mature protein suitable variant polypeptides have sequences that are at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, preferably at least about 95%, 96%, 97%, more preferably at least about 98%, or most preferably at least about 99% amino acid identity, that retain biological activity. Fragments of the proteins of the present invention which comprise at least 10 contiguous amino acids of a sequence disclosed herein and that are capable of exhibiting a biological activity of the corresponding full length protein are also encompassed by the present invention.

The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable

mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are capable of being fully secreted from the cell in which it is expressed.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

A "fragment" of a polypeptide is meant to refer to any portion of the molecule, such as the peptide core, a variant of the peptide core, or an extracellular region of the polypeptide. A "variant" of a polypeptide is meant to refer to a molecule substantially similar in structure and biological activity to either the entire molecule, or to a fragment thereof. Thus, provided that two molecules possess a similar activity, they are considered variants as that term is used herein even if the composition or secondary, tertiary, or quaternary structure of one of the molecules is not identical to that found in the other, or if the sequence of amino acid residues is not identical. An "analogue" of a polypeptide or genetic sequence is meant to refer to a protein or genetic sequence substantially similar in function and structure to the isolated polypeptide or genetic sequence.

It is understood herein that conservative amino acid substitutions can be performed to a purified and isolated polypeptide comprising any one of the

sequences of SEQ ID NOS.: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46 which are likely to result in a polypeptide that retains biological or immunological activity, especially if the number of such substitutions is small. By "conservative amino acid substitution" is meant substitution of an amino acid with an amino acid having a side chain of a similar chemical character. Similar amino acids for making conservative substitutions include those having an acidic side chain (glutamic acid, aspartic acid); a basic side chain (arginine, lysine, histidine); a polar amide side chain (glutamine, asparagine); a hydrophobic, aliphatic side chain (leucine, isoleucine, valine, alanine, glycine); an aromatic side chain (phenylalanine, tryptophan, tyrosine); a small side chain (glycine, alanine, serine, threonine, methionine); or an aliphatic hydroxyl side chain (serine, threonine).

Microarrays

Another aspect of the invention is a composition comprising a plurality of polynucleotide probes for use in detecting gene expression pattern(s) characteristic of particular cell type(s) and for detecting changes in the expression pattern of a particular cell type, e.g., lymphatic endothelial cells. For example, the invention comprehends an array, such as a microarray, comprising polynucleotides having at least 10 contiguous nucleotides selected from the polynucleotide sequences in the sequence listing, or a gene found in any of Tables 1 and 2 herein or Tables 3 and 4 of PCT/US03/06900.

Also contemplated are microarrays comprising polynucleotides having at least 10 contiguous nucleotides selected from the group of SEQ ID NOS:. 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 as well as any one of the genes found Tables 3 and 4 of PCT/US03/06900. Microarrays of the invention comprise at least 3 polynucleotides, wherein each enumerated polynucleotide has a distinct sequence selected from the group consisting of SEQ ID NOS:. 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 and a genbe found in any of Tables 3 and 4 of PCT/US03/06900. Such microarrays may also have duplicate polynucleotides and additional polynucleotides, e.g., control polynucleotides for use in hybridization-based assays using the microarray. Arrays, including microarrays,

having more than three distinct polynucleotides according to the invention, such as at least five, seven, nine, 20, 50 or more such polynucleotides, will be recognized as arrays according to the invention having the capability of yielding subtle distinctions between biological samples such as various endothelial cell types, or of providing a different, and typically greater, level of confidence in the various uses of such arrays, e.g., in screening for particular endothelial cells, in screening for abnormal or diseases cells and tissues, and the like.

The term "microarray" refers to an ordered arrangement of hybridizable array elements. The array elements are arranged so that there are preferably at least three or more different array elements, more preferably at least 100 array elements, and most preferably at least 1,000 array elements, on a solid support. Preferably, the solid support is a 1 cm^2 substrate surface, bead, paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. The hybridization signal from each of the array elements is individually distinguishable. In a preferred embodiment, the array elements comprise polynucleotide probes.

Hybridization means contacting two or more nucleic acids under conditions suitable for base pairing. Hybridization includes interaction between partially or perfectly complementary nucleic acids. Suitable hybridization conditions are well known to those of skill in the art. In certain applications, it is appreciated that lower stringency conditions may be required. Under these conditions, hybridization may occur even though the sequences of the interacting strands are not perfectly complementary, being mismatched at one or more positions. Conditions may be rendered less stringent by adjusting conditions in accordance with the knowledge in the art, e.g., increasing salt concentration and/or decreasing temperature. Suitable hybridization conditions are those conditions that allow the detection of gene expression from identifiable expression units such as genes. Preferred hybridization conditions are stringent hybridization conditions, such as hybridization at 42°C in a solution (*i.e.*, a hybridization solution) comprising 50% formamide, 1% SDS, 1 M NaCl, 10% dextran sulfate, and washing for 30 minutes at 65°C in a wash solution comprising 1 X SSC and 0.1% SDS. It is understood in the art that conditions of equivalent stringency can be achieved through variation of

temperature and buffer, or salt concentration, as described in Ausubel, *et al.* (Eds.), Protocols in Molecular Biology, John Wiley & Sons (1994), pp. 6.0.3 to 6.4.10. Modifications in hybridization conditions can be empirically determined or precisely calculated based on the length and the percentage of guanosine/cytosine (GC) base pairing of the probe. The hybridization conditions can be calculated as described in Sambrook, *et al.*, (Eds.), Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York (2d. Ed.; 1989), pp. 9.47 to 9.51.

One method of using probes and primers of the invention is in the detection of gene expression in human cells. Normally, the target will be expressed RNAs, although genomic DNA or a cDNA library may be screened. By varying the stringency of hybridization and the target binding site (i.e., the sequence of the probe, corresponding to a subset of one of the sequences set forth at SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 or a gene found in any of Tables 3 and 4 of PCT/US03/06900), different degrees of homology are expected to result in hybridization.

The microarray can be used for large-scale genetic or gene expression analysis of a large number of target polynucleotides. The microarray can also be used in the diagnosis of diseases and in the monitoring of treatments. Further, the microarray can be employed to investigate an individual's predisposition to a disease. Furthermore, the microarray can be employed to investigate cellular responses to infection, drug treatment, and the like.

The nucleic acid probes can be genomic DNA or cDNA or mRNA polynucleotides or oligonucleotides, or any RNA-like or DNA-like material, such as peptide nucleic acids, branched DNAs, and the like. The probes can be sense or antisense nucleotide probes. Where target polynucleotides are double-stranded, the probes may be either sense or antisense strands. Where the target polynucleotides are single-stranded, the probes are complementary single strands. In one embodiment, the probes are cDNAs. The size of the DNA sequence of interest may vary and is preferably from 100 to 10,000 nucleotides, more preferably from 150 to 3,500 nucleotides.

The probes can be prepared using a variety of synthetic or enzymatic techniques, which are well known in the art. The probes can be synthesized, in whole or in part, using chemical methods well known in the art (Caruthers et al., *Nucleic Acids Res., Symp. Ser.*, 215-233, 1980).

Pharmaceutical Formulations and Routes of Administration

A protein of the present invention (from whatever source derived, such as from recombinant and non-recombinant sources) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers, diluents, adjuvants or excipients at doses to treat or ameliorate a variety of disorders. Such a composition may also contain (in addition to protein and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, chemokines, lymphokines, growth factors, or other hematopoietic factors such as a PDGF, a VEGF (particularly a VEGF-C or a VEGF-D), VEGFR-3 (including soluble VEGFR-3 peptides comprising an extracellular domain), M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. Various forms of these polypeptides are contemplated as well, such as isolated holoproteins, subunits, fragments (*e.g.*, soluble fragments), and peptide fusions. The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with a protein of the invention, or to minimize side effects. Conversely, a protein of the present invention may be included in formulations of the particular cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the cytokine,

lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent. A protein of the present invention may be active in multimers (*e.g.*, heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, Pa., latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of beneficial change, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing methods of treatment or use of the present invention, a therapeutically effective amount of protein of the present invention is administered to a mammal having a condition or disorder to be treated. Protein of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, a protein of the invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering a protein of the invention in combination with a cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

Routes of Administration

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein of the present invention used in the pharmaceutical composition or to practice the method of the present invention is carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to a mammal, such as a human patient, is preferred. Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound at the site of intended action.

Compositions/Formulations

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein of the present invention is administered orally, protein of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein of the present invention, and preferably from about 25 to 90% protein of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid

form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein of the present invention, and preferably from about 1 to 50% protein of the present invention.

When a therapeutically effective amount of protein of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained by combination with a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose,

mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous

infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long-acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a cosolvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The cosolvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of

VPD diluted 1:1 with a 5% dextrose-in-water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.,* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compound over a time period of a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel-phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the proteinase-inhibiting compounds of the invention may be provided as salts with pharmaceutically compatible counterions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine, and the like.

The pharmaceutical compositions of the invention may be in the form of a complex of a protein(s) of the present invention along with protein or peptide antigens. The pharmaceutical compositions of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithin, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Pat. Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, each of which is incorporated herein by reference.

The amount of protein of the invention in the pharmaceutical composition will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein of the present invention and observe the patient's response. Larger doses of protein of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the methods of the invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein of the invention per kg body weight. When administered, the therapeutic composition for use in this invention is in a pyrogen-free, physiologically acceptable form. Therapeutically useful agents other than a protein of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either *in vivo* or *ex vivo* into cells for

expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

Effective Dosage

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve an intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of, or to alleviate the existing symptoms of, the subject being treated. Suitable properties that may be used in determining effective dosages include measurements of LEC and/or BEC growth stimulation or inhibition, rates or extent of cell differentiation into LECs and/or BECs, tendencies of cell expression patterns to shift towards or away from LEC- or BEC-specific expression patterns, and the like. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in a method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, for inhibitory methods, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibitory concentration). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or, in the case of life-threatening conditions, a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the

population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

Packaging

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

In addition, the invention comprehends a use of such a composition to manufacture a medicament for the treatment of a cell or an organism, such as a human patient, having a hyperproliferative or hypoproliferative disorder of a LEC and/or BEC, such as lymphedema, lymphangioma, lymphangiomyeloma, lymphangiomatosis, lymphangiectasis, lymphosarcoma, or lymphangiosclerosis, comprising administering an effective amount, or dose, of a composition according to the invention to the cell or organism. Suitable compositions include, but are not limited to, any polynucleotide according to the invention (e.g., an antisense polynucleotide), any polypeptide according to the invention, an antibody specifically

recognizing a polynucleotide or polypeptide according to the invention, a small molecule compound effective in modulating the expression of a polynucleotide according to the invention, and the like. Also contemplated are uses of compositions according to the invention for the manufacture of a medicament to ameliorate a symptom associated with a LEC- or BEC-associated disease or disorder.

Antibodies

Antibodies are useful for modulating the polypeptides of the invention due to the ability to easily generate antibodies with relative specificity, and due to the continued improvements in technologies for adopting antibodies to human therapy. Thus, the invention contemplates use of antibodies (e.g., monoclonal and polyclonal antibodies, single chain antibodies, chimeric antibodies, bifunctional/bispecific antibodies, humanized antibodies, human antibodies, domain antibodies and complementary determining region (CDR)-grafted antibodies, including compounds which include CDR sequences which specifically recognize a polypeptide of the invention), specific for polypeptides of interest to the invention. Preferred antibodies are human antibodies, such as those produced in transgenic animals, which are produced and identified according to methods described in WO 93/11236, published June 20, 1993, which is incorporated herein by reference in its entirety. Antibody fragments, including Fab, Fab′, F(ab′)2, and Fv, are also provided by the invention. The term "specific for," when used to describe antibodies of the invention, indicates that the variable regions of the antibodies of the invention recognize and bind the polypeptide of interest at a detectably different, and greater, level that bind to other substances (i.e., able to distinguish the polypeptides of interest from other known polypeptides of the same family, by virtue of measurable differences in binding affinity, despite the possible existence of localized sequence identity, homology, or similarity between family members). It will be understood that specific antibodies may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine binding specificity of an antibody of the

invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds.), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6.

Non-human antibodies may be humanized by any method known in the art. A preferred "humanized antibody" has a human constant region, while the variable region, or at least a complementarity-determining region (CDR), of the antibody is derived from a non-human species. Methods for humanizing non-human antibodies are well known in the art. See U.S. Patent Nos. 5,585,089, and 5,693,762. Generally, a humanized antibody has one or more amino acid residues introduced into its framework region from a source which is non-human. Humanization can be performed, for example, using methods described in Jones *et al.* (*Nature* 321: 522-525, 1986), Riechmann *et al.*, (*Nature*, 332: 323-327, 1988) and Verhoeyen *et al.* [*Science* 239:1534-1536, (1988)], by substituting at least a portion of a rodent CDR for the corresponding regions of a human antibody. Numerous techniques for preparing engineered antibodies are described, *e.g.*, in Owens and Young, (*J. Immunol. Meth., 168*:149-165, 1994). Further changes can then be introduced into the antibody framework to modulate affinity or immunogenicity.

The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Polypeptides and/or polynucleotides of the invention may also be used to immunize animals to obtain polyclonal and monoclonal antibodies which specifically react with the polypeptide. Such antibodies may be obtained using either the entire polypeptide or fragments thereof as an immunogen. The peptide immunogens additionally may contain a cysteine residue at the carboxyl terminus, and may be conjugated to a hapten such as keyhole limpet hemocyanin (KLH). Methods for synthesizing such peptides are known in the art, for example, as taught in R. P. Merrifield, *J. Amer. Chem. Soc. 85*:2149-2154 (1963); J. L. Krstenansky, et al., *FEBS Lett.* 211: 10 (1987). Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the polypeptide. Neutralizing monoclonal antibodies binding to the polypeptide may also be useful

therapeutics for both conditions associated with the polypeptide and also in the treatment of some forms of cancer where abnormal expression of the polypeptide is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the polypeptide are useful in detecting and preventing the metastatic spread of the cancerous cells mediated by the polypeptide. In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., Monoclonal Antibodies Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth et al., *J. Immunol. 35*:1-21 (1990); Kohler and Milstein, *Nature 256*:495-497 (1975), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., *Immunology Today 4*:72, 1983); Cole et al., in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985), pp. 77-96).

Any animal (mouse, rabbit, and the like) which is known to produce antibodies can be immunized with a peptide or polypeptide of the invention. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the polypeptide encoded by an ORF of the invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection. The protein that is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling the antigen with a heterologous protein (such as a globulin or β-galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal-antibody-producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, Western blot analysis, or radioimmunoassay

(Lutz *et al., Exp. Cell Research.* 175:109-124, 1988). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)). Techniques described for the production of single-chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce single-chain antibodies to polypeptide of the invention.

For polyclonal antibodies, antibody-containing antiserum is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The invention further provides the above-described antibodies in detectably labeled form. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, and the like) fluorescent labels (such as FITC or rhodamine, and the like), paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art; for example, see Sternberger, et al., *J. Histochem. Cytochem. 18*:315, 1970; Bayer,et al., *Meth. Enzym. 62*:308, 1979; Engval, et al., *Immunol. 109*:129, 1972; and Goding, *J. Immunol. Meth. 13*:215, 1976.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, and acrylic resins such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present

invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Computer-Readable Sequences

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer-readable media. As used herein, "computer-readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to, magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer-readable media can be used to create a manufacture comprising computer-readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer-readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer-readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer-readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention. By providing the nucleotide sequence of any of

SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 or a representative fragment thereof, or a nucleotide sequence at least 99.9% identical to any of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 in computer-readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer-readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., *J. Mol. Biol. 215*:403-410, 1990) and BLAZE (Brutlag et al., *Comp. Chem. 17*:203-207, 1993) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein-encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are

disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

Diagnostic Assays and Kits

The present invention further provides diagnostic assays, and related kits, for hyper- and/or hypo-proliferative disorders or diseases of endothelial cells such as LECs or BECs. These assays comprise methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or an antibody according to the invention.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex

with, the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample.

Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample. In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, Fla. Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described

method will vary based on the assay format, nature of the detection method and the tissues, and cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. In one embodiment, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibody or antibodies used in the assay, containers which contain wash reagents (such as phosphate-buffered saline, Tris buffers, and the like), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

EXAMPLES

Each reference cited in the methods described below is incorporated by reference herein. Methods used in the examples are as follows:

Antibodies

Monoclonal antibodies against human VEGFR-3 (clone 2E11D11; see International Patent Application No. PCT/US02/22164, published as WO 03/006104), PAL-E (Monosan), CD31 (Dako), N-cadherin, VE-cadherin, ß-catenin and plakoglobin and polyclonal rabbit anti-human podoplanin were used (Breiteneder-Geleff, S., et al., *Am. J. Pathol. 154*:385-394, 1999). Mouse anti-human integrin α9 was provided by Dr. Dean Sheppard (University of California at San Francisco, San Francisco) and Dr. Curzio Rüegg (University of Lausanne Medical School, Lausanne, Switzerland). The fluorochrome-conjugated secondary antibodies were obtained from Jackson Immunoresearch.

Cell Culture and Transfection

Human amniotic epithelial cells are cultured in Med199 medium in the presence of 5% fetal calf serum. Human dermal microvascular endothelial cells are obtained from PromoCell (Heidelberg, Germany). Anti-Podoplanin antibodies, MACS colloidal super-paramagnetic MicroBeads conjugated to goat anti-rabbit IgG antibodies (Miltenyi Biotech, Bergisch Gladbach, Germany), LD and MS separation columns and Midi/MiniMACS separators (Miltenyi Biotech) are used for cell sorting according to the instructions of the manufacturer. The isolated cells are cultured on fibronectin-coated (10 µg/ml, Sigma, St. Louis, MO) plates as described (Mäkinen, T., *et al.*, *EMBO J. 20*:4762-4773, 2001).

Immunofluorescence and Immunohistochemistry

To carry out immunofluorescence on isolated epithelial cells, the cells are cultured on coverslips, fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton-X100 in phosphate-buffered saline (PBS) and stained with primary antibodies. For integrin α9, staining live cells were incubated with the antibody for 15 minutes on

ice before fixation. The cells were further stained with FITC- or TRITC-conjugated secondary antibodies. F-actin is stained using TexasRed-conjugated phalloidin (Molecular Probes, Eugene, OR). Cells are counterstained with Hoechst 33258 fluorochrome (Sigma) and viewed using a Zeiss Axioplan 2 fluorescent microscope.

Normal human skin obtained after surgical removal is embedded in Tissue-Tek® (Sakura, The Netherlands), frozen and sectioned. The sections (6 μm) are fixed in cold acetone for 10 minutes and stained with the primary antibodies followed by peroxidase staining using Vectastain Elite ABC kit (Vector Laboratories, Burlingame, CA) and 3-amino-9-ethyl carbazole (Sigma, St. Louis, MO).

RNA isolation, Northern blotting and microarray analyses

Total RNA was isolated and DNAseI treated in RNeasy columns (Qiagen, Valencia, CA). ^{32}P-labeled probes for hybridization with the Atlas filters (Clontech) were prepared using 2-5 μg of total RNA according to the manufacturer's instructions with the exception that the probe was purified using Nick-25 columns (Pharmacia Biotech, Uppsala, Sweden). Following hybridizations and washes, the membranes were analyzed using a Fuji BAS 100 phosphoimager. For the Affymetrix® analysis, 2-5 RNA samples isolated from LEC and BEC cell cultures were hybridized separately using RNA extracted from four lots of cells isolated from different individuals. For the Affymetrix® expression analysis, 5 μg of total RNA was used for the synthesis of double-stranded cDNA using Custom SuperScript ds-cDNA Synthesis Kit (Invitrogen, Carlsbad, CA). Biotin-labeled cRNA was then prepared using the Enzo BioArrayTMHighYieldTMRNA Transcript Labeling Kit (Affymetrix, Santa Clara, CA), and the unincorporated nucleotides were removed using RNeasy columns (Qiagen, Valencia, CA). The hybridization, washing and staining of Human Genome U133A microarrays were done according to the instructions of the manufacturer (Affymetrix, GeneChip Expression Analysis Technical Manual). The probe arrays were scanned at 570 nm using an Agilent GeneArray® Scanner and the readings from the quantitative scanning were analyzed by the Affymetrix® Microarray Suite version 5.0 and Data Mining Tool version 3.0.

For the comparison analyses, the hybridization intensities were calculated using a global scaling intensity of 100.

The differentially expressed sequences were used for searching EST contigs in the GenBank database of the National Center for Biotechnology Information and the National Library of Medicine. (NCBI/NLM), and open reading frames were predicted using the orf finder software available at NCBI/NLM. The SOSUI system was used for prediction of transmembrane helices and signal sequences from the protein sequences, and other protein domain architectures were analyzed using Pfam (Protein families database of alignments and HMMs).

It is expected that the amino acid sequence of a given polypeptide, whether obtained from the Swissprot database, which provides the sequences for typical Pfam analyses, or a Genbank sequence, would be substantially, if not exactly, identical over a typically predominating overlap between those sequences. Moreover, the polypeptides characterized by either of the sequences would be expected to have the same or substantially similar biological activities. Differences in the two sequences may arise due to the fact that some polypeptides identified and used in the Pfam analyses may contain domains designated as "low complexity" or "coiled coil" that are not included in the sequence of a final polypeptide product, as provided by Genbank. Additionally, the Genbank polynucleotide sequences provided herein, which encode the polypeptides disclosed herein, are expected to correspond to polynucleotides encoding the polypeptides utilized in the Pfam analysis.

EXAMPLE 1

IDENTIFICATION OF CELL SURFACE PROTEINS EXPRESSED DIFFERENTIALLY IN LEC

Lymphatic endothelial cells (LEC) were isolated from cultures of human dermal microvascular endothelial cells using magnetic microbeads and antibodies against the lymphatic endothelial cell surface marker podoplanin (Breiteneder-Geleff, et al., *Am. J. Pathol. 154*:385-394 (1999); Mäkinen, et al., *EMBO J. 20*:4762-4773 (2001)), as described in PCT/US03/06900. The purities of the isolated LEC populations were confirmed to be over 99% as assessed by immunofluorescence using antibodies against VEGFR-3 or podoplanin. The isolated

cells were cultured for a couple of passages, and RNA was extracted from the cultures and used for hybridization with oligonucleotide microarrays containing sequences from about 12,000 known genes, *i.e.*, approximately 1/3 of the total number of all predicted human transcripts.

PCT/US03/06900, which examined LEC specific genes based on the 95Av2 and 9513-E arrays, identified approximately 165 LEC specific genes in several different protein families, such as adhesion molecules, cytoskeletal proteins, extracellular matrix proteins transcription factors and growth factors.

Experiments were performed to characterize LEC gene expression profiles revealed by the U133A microarray. The Human Genome U133 (HG-U133) Set, consisting of two GeneChip® arrays, contains almost 45,000 probe sets representing more than 39,000 human gene transcripts derived from approximately 33,000 well-substantiated human genes. This set design uses sequences selected from GenBank®, dbEST, and RefSeq. The HG-U133A Array includes representation of the RefSeq database sequences and probe sets. The HG-U133B Array contains primarily probe sets representing EST clusters. The U133A array utilized herein provides a greater sampling of human genes and ESTs (approximately 20,0000 more genes) than the arrays described previously in PCT/US03/06900.

Differentially expressed genes were delineated from the U133A array if their signal ratio was over 1.1 and p-value below 0.05, resulting in over 600 genes found to be differentially expressed between LECs and BECs.

LEC-specific genes were further analyzed using a subtraction library involving LEC and BEC genes. To construct the library, total RNA was isolated as previously described and 5μg of total RNA was pre-amplified using a SMART™ PCR cDNA synthesis kit (BD Biosciences Clontech). After RsaI-digestion, PCR-Select cDNA subtraction was carried out in both directions, resulting in selective amplification of differentially expressed sequences, and subtracted LEC and BEC cDNA libraries were prepared (BD Biosciences Clontech). Subtractive hybridization was performed with a 1 (tester): 30 (driver) ratio in both directions and subtracted cDNA pools were amplified by PCR. Forty nanograms of the purified PCR-amplified product were cloned into the pAtlas vector (PUC-based vector) for the construction of

subtracted libraries, although a number of other vectors could be used in the construction, as would be known in the art.

Differential screening of the subtracted LEC-specific library was carried out as described in the PCR-Select Differential Screening Kit User Manual (BD Biosciences, Clontech). The LEC-specific subtracted library was plated and individual bacterial clones were picked and grown. After DNA extraction, the inserts were amplified by PCR and used for sequencing. An aliquot of each PCR-amplified insert was also arrayed onto a nylon membrane and used for hybridization with ^{32}P-labeled cDNA probes. The results from the hybridizations with subtracted LEC-specific (tester) and subtracted BEC-specific (driver) cDNA probes were used for the differential expression analyses.

BLAST (The Basic Local Alignment Search Tool) was used to compare the sequences against nucleotide, protein and EST sequence databases. For unknown sequences, EST contigs were searched and open reading frames were predicted using ORF finder. Protein domain architectures were analyzed using Pfam (Protein families database of alignments and HMMs) and Smart (Simple Modular Architecture Research Tool).

The nucleotide sequences of clones that were differentially expressed in LECs versus BECs were analyzed in the manner described above. Several of the EST or unknown gene fragments detected in the first screen have been investigated further to determine their sequence similarities to known gene sequences and to identify any open reading frames and functional domain similarities.

Several of the LEC-specific genes have been found to correspond to KIAA gene sequences, which are large nucleotide EST clones encoding unknown human proteins. (Kazusa DNA Research Institute, 1532-3, Yana Kisarazu, Chiba, 292-0812, Japan). These LEC-specific genes were further analyzed in several available databases to determine the existence of species homologs and the percent similarity in these homologs and also to reveal amino acid sequences that demonstrate similarity to conserved protein domains.

Analyses of the LEC clone sequences were performed using the HomoloGene database maintained by the U.S. National Center for Biotechnology

Information offered by the National Institutes of Health to determine species homologs and orthologs and their percent similarity to the newly isolated human LEC-specific genes. Analyses of the sequences were performed using a resource of curated and calculated homologs for genes as represented by UniGene or by annotation of genomic sequences, generally comparing EST and mRNA sequences from UniGene, as well as transcripts extracted from annotated genomic sequences. (Zhang, *et al., J. Comp. Biol.* 7:203-14. 2000). The best match for a nucleotide sequence in one organism to a nucleotide sequence in a second organism is based on the degree of similarity between the two sequences, with a minimum alignment of 100 base pairs. The similarity between the two sequences was determined by an alignment score. The alignment score for a sequence pair is the sum of the similarity scores of the sections of the two sequences that aligned.

BLAST and Pfam analysis of the isolated genes revealed numerous LEC-specific transmembrane proteins, including KIAA1972, KIAA0472, KIAA0494, ATP10D, KIAA1724, DKFZp564K142, KIAA1573, DKFZp761N1114, and FLJ25916, set out in SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43 (see Table 1).

Table 1

LEC-Specific Transmembrane Proteins

Protein	Genbank Accession Number	nucleotide SEQ ID NO:	polypeptide SEQ ID NO:
KIAA1972	XM_166279	SEQ ID NO: 5	SEQ ID NO: 6
KIAA0472	XM_290898	SEQ ID NO: 7	SEQ ID NO: 8
KIAA0494	NM_014774	SEQ ID NO: 9	SEQ ID NO: 10
ATP10D	NM_020453	SEQ ID NO: 11	SEQ ID NO: 12
KIAA1724	XM_040280	SEQ ID NO: 13	SEQ ID NO: 14
DKFZp564K142	NM_032121	SEQ ID NO: 15	SEQ ID NO: 16
KIAA1573	NM_020925	SEQ ID NO: 19	SEQ ID NO: 20
DKFZp761N1114	AI332934/ NM_181644	SEQ ID NO: 21	SEQ ID NO: 22

FLJ25916	AK098782/ NM_207347	SEQ ID NO: 23	SEQ ID NO: 24

KIAA1972: LEC-specific gene KIAA1972 (Genbank Accession No. XM_166279; SEQ ID NOS: 5 and 6), a 3540 bp mRNA, has been detected by Northern blot expression at high levels in liver, kidney, testis, spleen and ovary, at moderate levels in heart, lung, smooth muscle, pancreas, spinal cord, and fetal liver, and at low levels in brain and fetal brain. Northern blot data was derived from the electronic database of Human Unidentified Gene-Encoded (HUGE) Large Proteins maintained by Kazusa DNA Research Institute, Chiba, Japan (see Ohara et al., *DNA Res.* 4:53-59, 1997).

Unigene comparison of the KIAA1972 sequence indicates the protein is 36% identical [over a string of 96 amino acids (aa)] to human Ryanodine receptor 2, RYR2, which is a cardiac muscle-type ryanodine receptor calcium release channel. KIAA1972 also shows 34% identity (over a string of 110 aa) to a fragment of mouse (*Mus musculus*) ryanodine receptor, demonstrates 29% identity (over a sequence of 152 aa) to a *Drosophila melanogaster* ryanodine receptor/calcium release channel, shows 48% identity (over a string of 497 aa) to a *Caenorhabditis elegans* zinc finger, C3HC4 type (RING finger) protein, and shows 39% identity (over 140 aa) to an *Arabidopsis thaliana* hypothetical protein. KIAA1972 also demonstrates similarity to species orthologs *R. norvegicus* LOC291860 (92.4% similarity) and *M.musculus* 4930470D19Rik (87% similarity).

Protein domain mapping for the KIAA1972 sequence using the Pfam database (SwissProt designation Q96DX4) predicts a protein of 576 amino acids with a signal peptide from amino acids 1-16, a SPRY domain from approximately residues 358-482 and a RING domain from approximately residues 527-561. The SPRY domain is derived from similarities to the spla motif and the ryanodine receptor. The SPRY domains have homologues in butyrophilin/marenostrin/pyrin proteins. Ca^{2+}-release from the sarcoplasmic or endoplasmic reticulum, the site of intracellular Ca^{2+} storage, is mediated by the ryanodine receptor (RyR) and/or the inositol trisphosphate receptor (IP3R).

The RING-finger domain is a specialized type of Zn-finger region of 40 to 60 amino acid residues that binds two zinc atoms, and is probably involved in mediating protein-protein interactions. RING domains contain cysteine and histidine residues for binding to the Zinc molecules in either a C3HC4-or a C3H2C3 configuration. E3 ubiquitin-protein ligase activity is intrinsic to the RING domain of c-Cbl and is likely to be a general function of the RING domain. E3 ubiquitin-protein ligase MdM2 expression is amplified in certain tumors (including soft tissue sarcomas, osteosarcomas and gliomas), indicating that ubiquitin-ligase function may be important in maintaining cell homeostasis and regulating cell cycle activities.

Therefore, KIAA1972 is expected to act to regulate calcium storage in lymphatic endothelial cells and to function in the regulation of cell homeostasis. Deregulation of KIAA1972 is expected to result in aberrant LEC proliferation and lymphatic disorders such as lymphedema, lymphangioma, lymphangiomyeloma, lymphangiomatosis, lymphangiectasis, lymphosarcoma, and lymphangiosclerosis.

KIAA0472: Full-length KIAA0472, first isolated from the human adult brain genome, is a 2589 bp mRNA sequence. Unigene comparison of the KIAA0472 gene (Genbank Accession No. XM_290898; SEQ ID NOS: 7 and 8) indicates that the predicted protein is 65% identical (within a sequence of 102 aa) to human neuronal thread protein, shares 29% identity (over 245 aa) to *M. musculus* death domain-containing protein RIP; and demonstrates 33% identity (over 139 aa) with rat *R. norvegicus*) protein kinase C, delta type, NPKC-delta. KIAA0472 also demonstrates significant identity to several other protein kinases such as *D. melanogaster* TGF-beta activated kinase 1, DmTak1 (27% identity over 262 aa), a *C. elegans* tyrosine-protein kinase (25% identity over 293 aa) and an *A. thaliana* protein kinase (38% identity over 161 aa).

Protein domain mapping of the predicted 390-amino-acid KIAA0472 gene (SwissProt designation O75060) reveals that a 365 amino acid fragment of the polypeptide (SEQ ID NO: 42) possesses a protein kinase domain from approximately residues 88-342. Pfam analysis predicts the domain lies at residues 88-333 while SMART analysis predicts a STYKc domain, from approximately residues 88-342. There is also a region of similarilty to a potential transmembrane domain from amino acids 262-284. The designation of the STYKc domain leads to the expectation that

KIAA0472 is a protein kinase of unclassified specificity, and is more specifically expected to be a phosphotransferase exhibiting dual-specificity as a Ser/Thr/Tyr kinase. Both serine/threonine and tyrosine protein kinases share a common catalytic core comprising a glycine-rich region of amino acids in proximity to a lysine residue which has been shown to be involved in ATP binding, as well as a conserved aspartic acid residue which is important for the catalytic activity of the enzyme.

KIAA0494: KIAA0494 is a mRNA sequence of 5766 base pairs first isolated from human adult brain. Unigene comparison of the KIAA0494 sequence (Genbank Accession No. NM_014774; SEQ ID NOS: 9 and 10) indicates the predicted protein is 23% identical (over a sequence of 306 amino acids) to *Arabidopsis thaliana* hypothetical protein F5E19_70. Analysis of potential species orthologs revealed that KIAA0494 is similar to *R. norvegicus* LOC298425 (85.7% similarity) and *M.musculus* 4732418C07Rik (82.2% similarity).

Protein domain mapping using the Pfam database indicates that the 495-amino-acid KIAA0494 protein (SwissProt designation O75071) exhibits a transmembrane domain from approximately amino acids 71-93; and, coiled-coil regions from amino acids 95-115, 209-229 and 292-312. Both transmembrane and coiled coil domains are found in proteins with diverse functions. Presence of the transmembrane domains indicates the protein is membrane bound. Coiled-coil domains are structural domains composed of two or three alpha helices that wind around each other and are found in a variety of proteins.

ATP10D: Screening for differential expression of LEC- and BEC-specific genes identified the ATP10D gene (Genbank Accession No. NM_020453; SEQ ID NOS: 11 and 12) as a gene upregulated in lymphatic endothelia. ATP10D was isolated as a 6550 bp mRNA sequence. Analysis of related species orthologs detects an ATP10D ortholog in *R. norvegicus* LOC305314 (87.2 %).

Unigene comparison of the ATP10D gene sequence indicates that the predicted ATP10D protein demonstrates 99% identity, over a 650 aa length of aligned region, to human phospholipid-transporting ATPase VD (ATPVD); demonstrates 56% identity, over a 580 aa alignment, to a mouse potential phospholipid-transporting ATPase VA; demonstrates 23% identity (over a 365 amino acid alignment) to rat ATPase isoform 2, Na^+K^+ transporting, beta polypeptide 2; shows 38% identity (604

aa alignment) to an *A. thaliana* putative chromatin granule ATPase II homolog; and is 41% identical (over a 570 aa alignment) to a *C. elegans* ATPase.

Human ATP10D is predicted to be a 1426-amino-acid protein. Protein domain mapping of the 1416 amino acid mouse homolog of human ATP10D (SwissProt designation Q8K2X1) demonstrates five transmembrane domains, from amino acids 7-25, 53-75, 88-105, 115-137, and 157-176, and also to possess a signal peptide from residues 1-23. The mouse ATP10D also contains a hydrolase domain from amino acids 805-1074. Hydrolase enzymes catalyze the hydrolysis of various molecular bonds and include such proteins as epoxide hydrolases and phosphatases.

Human ATP10D also contains predicted regions of similarity to MgtA domains involved in cation transport, from approximately residues 70-457 and residues 722-1329 of SEQ ID NO: 12, and is expected to be a cation transport ATPase, similar to other ATPase transporters that are involved in inorganic ion transport and metabolism. Other possible functions include cation transport, integral to membrane, magnesium ion binding, hydrolase activity, ATPase activity, and phospholipid-translocating ATPase activity. ATP10D is expected to play a significant role in transporting ions across the LEC membrane and regulating varied cellular activities.

KIAA1724: KIAA1724 (Genbank Accession No. XM_040280; SEQ ID NOS: 13 and 14) was isolated as a LEC-specific, 8083 bp mRNA sequence. Tissue expression analysis by Northern blot (data derived from HUGE protein database maintained by Kazusa DNA Research Institute, Chiba, Japan) detected KIAA1724 transcripts at highest levels in B. amygdala, B. cerebellum, B. substantia nigra, and B. thalamus, with moderate expression in overall brain, liver, B.corpus callosum, B.caudate nucleus, B. hippocampus and the B. subthalamic nucleus. Analysis of species similarities led to the identification of species orthologs in *R. norvegicus* LOC298856 (88.4 %) and *M. musculus* 4933402G07Rik (84.4 %)

Unigene analysis of the KIAA1724 sequence reveals that the sequence demonstrates 64% identity, over a string of 49 amino acids, to a probable human thromboxane A2 receptor isoform beta, and 34% identity, over a 377 amino acid alignment, to *C. elegans* hypothetical protein F54D7.2.

Protein domain mapping of a 400-residue fragment comprising the predicted 386 residue KIAA1724 protein (SwissProt designation Q9COD9) indicates the protein contains a CDP-alcohol phosphatidyltransferase (CDP-OH P transf) domain from amino acids 108-252, and 10 transmembrane domains located at residues 61-83; 98-117; 137-159; 164-186; 193-215; 235-257; 270-292; 305-324; 331-53; and 358-380. Phosphatidyltransferases are involved in phospholipid biosynthesis. Generally, these enzymes are proteins of from 200 to 400 amino acid residues exhibiting a conserved region that contains three aspartic acid residues and is located in the N-terminal region of the proteins. KIAA1724 is expected to function in phospholipid biosynthesis in LECs.

DKFZp564K142: Analysis of the genome for LEC-specific genes identified hypothetical gene DKFZp564K142 (Genbank Accession No. NM_032121; SEQ ID NOS: 15 and 16), which comprises a nucleotide sequence of 2788 bp, and exhibits higher expression in LECs versus blood endothelial cells. Initial analysis of the sequence revealed DKFZp564K142 species orthologs in *S. scrofa* Ssc.16415 (93.9%), *R. norvegicus* IAG2 (87.8 %) and *M. musculus*, 2610529C04Rik (79.7 %).

Unigene analysis of DKFZp564K142 indicates that the amino acid sequence of the protein is 69% identical over a 323 aa region to the protein encoded by human gene N33; is 93% identical, over 332 aa, to the rat implantation-associated protein, IAG2; and is 42% identical (over a 303 aa sequence) to the *C. elegans* ZK686.3 protein.

Pfam analysis of the 335 amino acid hypothetical DKFZp564K142 protein (SwissProt designation Q9HOU3) indicates that the protein comprises an OST3-OST6 domain located from approximately residues 20-330. The proteins in the yeast OST3-OST6 family are part of a complex of eight endoplasmic reticulum (ER) proteins that transfer core oligosaccharide from dolichol carrier to Asn-X-Ser/Thr motifs. This family includes both OST3 and OST6, each of which contains four predicted transmembrane helices. Disruption of OST3 and OST6 leads to a defect in the assembly of the protein complex. Hence, the function of these genes is expected to be essential for recruiting a fully active complex necessary for efficient N-glycosylation.

The DKFZp564K142 protein also contains a signal peptide at amino acids 1-29 and multiple transmembrane domains, similar to either the OST3 or OST6 protein, which are located at approximately amino acids 7-29, 186-205, 212-234, 266-288, and 301-320. Both rat IAG2 and DKFZp564K142 exhibit five transmembrane domains. Based on the high similarity to rat IAG2, DKFZp564K142 is expected to be a human homolog to implantation-associated protein involved in pregnancy and may play a role in establishing lymphatic endothelial cells during pregnancy and embryo implantation.

KIAA1573: KIAA1573 (Genbank Accession No. NM_020925; SEQ ID NOS: 19 and 20) encodes a LEC-specific, hypothetical human protein and is approximately 5565 base pairs in length. Northern blot analysis of mRNA expression (data derived from HUGE protein database maintained by Kazusa DNA Research Institute, Chiba, Japan) shows that KIAA1573 is expressed at high levels in B. thalamus, with lower expression in other tissues, such as B. caudate nucleus, B. hippocampus, B. subthalamic nucleus, B.substantia nigra, ovary, brain and kidney. Comparison of KIAA1573 to other species leads to the expectation of orthologs such as *R. norvegicus* LOC298267 (90.6%), *B. taurus* Bt.1428 (88.6%) and *M. musculus* 1190007F10Rik (83.2%).

Unigene analysis of the predicted amino acid sequence of KIAA1573 reveals that the protein demonstrates 22% identity (over a 598 aa sequence) to human calcium channel, voltage-dependent, alpha 2 delta calcium channel subunit; demonstrates 22% identity (within a sequence of 676 aa) to mouse calcium channel, voltage-dependent, alpha 2 delta calcium channel subunit; and 19 % identity (over 677 aa) to rat CIC2_RAT Dihydropyridine-sensitive L-type, calcium channel alpha-2/delta subunit precursor.

Pfam protein domain analysis of a 1185 residue polypeptide sequence (SEQ ID NO: 43) comprising the predicted 985 amino acid KIAA1573 protein (Swissprot designation Q9HCJ9) reveals two Cache domains in KIAA1573, a full Cache domain from amino acids 364-443 and a partial domain from 683-764. The protein also possesses a single transmemebrane domain spanning residues 1006-1028 of SEQ ID NO: 43, and is expected to be expressed on the cell surface. The cache

motif is a signaling domain that is found in animal calcium channel subunits and a certain class of prokaryotic chemotaxis receptors.

Analysis of the protein sequence using the Smart [Simple Modular Architecture Research Tool] database predicts the presence of a VWA domain from residues 137-333. The VWA domain is derived from von Willebrand factor A, which is a large multimeric glycoprotein found in blood plasma. This vWF domain is found in various plasma proteins: complement factors B, C2, CR3 and CR4; the integrins (I-domains); collagen types VI, VII, XII and XIV; and other extracellular proteins. The majority of VWA-containing proteins are extracellular. A common feature of proteins with this domain appears to be involvement in multiprotein complexes. Proteins that incorporate vWF domains participate in numerous biological events (e.g. cell adhesion, migration, homing, pattern formation, and signal transduction), involving interaction with a large array of ligands. VWA domains in extracellular eukaryotic proteins mediate adhesion via metal ion-dependent adhesion sites (MIDAS). A number of human diseases arise from mutations in VWA domains, including diseases of the lymphatic endothelium, such as bleeding disorders. The similarity of KIAA1573 to extracellular, adhesion type proteins leads to the expectation that it acts as a mediator of cell signaling, while the presence of the VWA domain implicates KIAA1573 as a useful target in regulating LEC function in lymphatic endothelium mediated bleeding disorders.

DKFZp761N1114: DKFZp761N1114 (Genbank Accession No. AI332934, NM_181644; SEQ ID NOS: 21 and 22) has a nucleotide sequence of 3352 bp. Tissue expression analysis of human tissues, as detected by Multiple Tissue Northern Blot (Clontech) of adult human RNA, indicates that DKFZp761N1114 produces a 4.5 kb transcript, which is expressed at high levels in the stomach, kidney, and brain, with lower expression in other tissues, such as skeletal muscle, colon, placenta, lung, tongue and trachea. Comparison of the sequence with other species leads to expected orthologs in *R. norvegicus* LOC304787 (85.9 %) and *M. musculus* - A930031D07Rik (80.3 %).

Unigene analysis of the predicted protein sequence shows that DKFZp761N1114 is 60% identical (over a 100 aa sequence) to human neuronal thread protein.

Protein domain mapping reveals the 403 amino acid DKFZp761N1114 (SwisProt designation Q8N468) protein contains a signal sequence and 7 apparent transmembrane domains, from amino acids 27-49; 108-130; 194-216; 236-258; 265-287; 291-313; 326-348; and 353-375. DKFZp761N1114 may also contain a fucose permease domain from approximately residues 202-376, which is involved in carbohydrate transport and metabolism. As such, DKFZp761N1114 is expected to act as a unique seven-transmembrane receptor on the surface of LECs and to modulate or regulate carbohydrate activity in the cell.

Further analysis revealed that the DKFZp761N1114 gene encodes an alternative predicted protein structure which includes a transmembrane protein with only a single transmembrane domain and a MSF-1 domain (Major Facilitator Superfamily), a domain present for example in glycerol-3-phosphate transporter from *E. coli* bacteria. An additional alternative predicted domain structure indicates DKFZp761N1114 may encode twelve transmembrane domains and an amino-terminal signal peptide.

The gene is expressed in lymphatic endothelial cells (LECs), but not in blood vascular endothelial cells (BECs), in an *in vitro* cell culture of primary endothelial cells (HDMEC, PromoCell).

In vivo, the gene is expressed, as revealed by non-radioactive *in situ* hybridization (NR-ISH), in lymphatic vessels of mouse embryonic mesenterium at E17.5, but is not detectable in blood vessels. The antisense transcript, as detected by NR-ISH using the sense probe, is also expressed in the mesenteric lymphatic vessels, but to a lesser extent. Thus, the expression of DKFZp761N1114 is expected to be modulated at least in part, by its antisense transcript. Using frozen sections from E16.5 mouse embryos, very little expression in embryonic mouse tissues was detected.

A Multiple Tissue Northern Blot (Clontech) of adult mouse RNA, indicates the DKFZp761N1114 gene is expressed in kidney and to a lesser extent in liver, brain and heart, but not in lung, testis, spleen or skeletal muscle. NR-ISH analysis showed that the gene was expressed in collecting tubules in the kidney of adult mice.

The expression of DKFZp761N1114 in kidney tubules and its domain structure, involving a Major Facilitator Superfamily (MFS) domain, indicate a function in uptake of substances from the extracellular space in kidney, as well as in lymphatic vessels.

FLJ25916: Screening of the LEC library also identified hypothetical protein FLJ25916 (Genbank Accession No. AK098782; SEQ ID NOS: 23 and 24) as a LEC-specific gene. FLJ25916 was isolated from fis, clone CBR04903 as a 1893 bp fragment.

Comparison of the putative coding region of the FLJ25916 mRNA to other sequences reveals that FLJ25916 is similar to human hypothetical protein KIAA0233 (44% identity over a 572 amino acid sequence) and is 28% identical (over 315 aa) to *C. elegans* protein T20D3.11.p. KIAA0233, to which FLJ25916 shows significant identity, demonstrates homology to both the mouse collagen alpha chain precursor and rat procollagen (26% identity with both proteins, over a sequence of approximately 1005 aa). Ortholog analysis for FLJ 25916 predicts similarity to mouse protein ENSMUSP00000059624 (novel Ensembl prediction).

Protein domain mapping of the hypothetical 583 amino acid FLJ25916 protein (SwissProt designation Q8N787) identified a signal peptide from residues 1-18, nine transmembrane domains located at approximately amino acids 2-19; 50-72; 220-242; 247-264; 269-291; 340-362; 401-423; 427-444; and 457-479, and also predicted a coiled-coil region from amino acids 169-189. The similarity of FLJ25916 to other mammalian collagen-like proteins leads to the expectation that it acts as a structural protein and/or interacts with the extracellular matrix surrounding the LEC.

Further analysis shows that FLJ25916, also identified as C18orf30, is expressed in lymphatic endothelial cells (LECs) and to a lesser extent in blood vascular endothelial cells (BECs) in an *in vitro* cell culture of primary endothelial cells (HDMEC, PromoCell). *In vivo*, the gene is expressed in the trigeminal ganglion in the brain, in the eye and in bone and is expressed weakly in submandibular gland, kidney and lung, as revealed by non-radioactive *in situ* hybridization (NR-ISH) of frozen sections from E16.5 mouse embryos.

The above transmembrane and cell-surface-expressed proteins are useful for the identification and isolation of LEC, and useful for regulation of LEC activity.

EXAMPLE 2

IDENTIFICATION OF INTRACELLULAR PROTEINS EXPRESSED DIFFERENTIALLY IN LEC

Assays using the microarray and LEC cDNA libraries identified several LEC-specific genes encoding proteins expressed on the cell surface and which are suitable targets for identification and localization of lymphatic endothelial cells, as well as being suitable targets for the manipulation of their function and activity. In addition to these cell-surface LEC proteins, intracellular proteins were identified as being LEC-specific; these proteins are expected to contribute to the regulation of the intracellular function of LEC genes, such as the transcription of particular genes and the transport of necessary proteins through the intracellular space.

By way of example, the analysis of LEC-specific intracellular proteins identified the following genes as being expressed at higher levels in LECs versus BECs: HERC3, KIAA1340, protein FLJ20051, protein FLJ13910, protein FLJ11029, KIAA1102, protein FLJ32029, protein FLJ23594, and protein FLJ10532 (see Table 2).

Table 2

LEC-Specific Transcription Factors or Intracellular Proteins

Protein	Genbank Accession Number	nucleotide SEQ ID NO:	polypeptide SEQ ID NO:
HERC3	NM_014606	SEQ ID NO: 27	SEQ ID NO: 28
KIAA1340	XM_044836	SEQ ID NO: 29	SEQ ID NO: 30
FLJ20051	NM_019087	SEQ ID NO: 31	SEQ ID NO: 32
FLJ13910	NM_022780	SEQ ID NO: 17	SEQ ID NO: 18
FLJ11029	NM_018304	SEQ ID NO: 33	SEQ ID NO: 34

KIAA1102	XM_044461	SEQ ID NO: 35	SEQ ID NO: 36
FLJ32029	NM_173582	SEQ ID NO: 37	SEQ ID NO: 38
FLJ23594	NM_024781	SEQ ID NO: 39	SEQ ID NO: 40
FLJ10532	AK001394	SEQ ID NO: 41	
BIKE protein	NM_017593	SEQ ID NO: 25	SEQ ID NO: 26

HERC3: Screening for LEC-specific genes identified the HERC3 gene as a gene upregulated preferentially in LECs versus BECs. Unigene comparison of the HERC3 sequence (Genbank Accession No. NM_014606; SEQ ID NOS: 27 and 28) indicates that the predicted protein is 95% identical (over 891 aa) to human hypothetical protein 2 (fragment); 35% identical (over 337 aa) to mouse rjs protein; shares 31% identity (over 184 aa) to rat ubiquitin ligase Nedd4; demonstrates 37% identity (over 309 aa) to the *A. thaliana* UVB-resistance protein, UVR8; is 24% identical (over a 330 aa sequence) to *D. melanogaster* chromatin-binding protein BJ1; and is 30% identical (over 891 aa) to *C .elegans* hypothetical protein Y48G8AL.1.p.

HERC3 protein domain mapping (SwissProt designation Q15034) reveals four regulator-of-chromosome-condensation (RCC1)-like protein domains found from residues 52-99, 102-152, 155-205, and 208-257. Regulator-of-chromosome-condensation (RCC1) is a eukaryotic protein that binds to chromatin and interacts with ran, a nuclear GTP-binding protein, to promote the loss of bound GDP and the uptake of fresh GTP, thus acting as a guanine-nucleotide dissociation stimulator (GDS). The interaction of RCC1 with ran is believed to play an important role in the regulation of gene expression. RCC1 is a protein that contains seven tandem repeats of a domain of about 50 to 60 amino acids. The RCC1-type of repeat is also found in the X-linked retinitis pigmentosa GTPase regulator.

Mapping of the HERC3 protein also reveals a HECT domain spanning approximately amino acids 721-1050, Pfam predicting the domain from amino acids 752-1050 and SMART analysis detecting the domain at residues 721-1050. HECT (Homologous to the E6-AP Carboxyl Terminus) domains are typically involved in ubiquitin ligase activity. Proteins containing this domain at the C-terminus include

ubiquitin-protein ligase, which regulates ubiquitination of CDC25, and directly transfers ubiquitin to targeted substrates. Human thyroid receptor interacting protein 12 also contains this domain and is a component of an ATP-dependent multi-subunit protein that interacts with the ligand-binding domain of the thyroid hormone receptor. This protein may have similar function to an E3 ubiquitin-protein ligase such as human ubiquitin-protein ligase E3A, which interacts with the E6 protein of the cancer-associated human papillomavirus types 16 and 18. A cysteine residue in the domain is required for ubiquitin-thiolester formation. The HECT domain is known to form complexes with the ubiquitin-conjugating domain of other proteins, and this domain is involved in ubiquinating proteins to signal their degradation by the proteosome.

Further analysis of protein domains in HERC3 indicate that a region from residues 333-481 is typically associated with proteins containing the classical zinc finger C2H2 domain; and residues 482-605 represent a domain associated with proteins containing lectin c domains.

Based on the foregoing information, HERC3 is expected to function as a regulator of protein degradation in LECs, given the similarity to ubiquitin-ligase proteins. HERC3 is also expected to be involved in chromatin binding, based on the presence of the RCC1 domain.

KIAA1340: Northern blot expression of KIAA1340 (Genbank Accession No. XM_044836; SEQ ID NOS: 29 and 30) indicates that KIAA1340 mRNA (approximately 4363 bp) is expressed at the highest levels in brain and at moderate levels in lung, ovary and spinal cord. *In situ* expression measured in day 16.5 mouse embryo indicates that the transcript is expressed at the highest level in the thymus and the intestine of developing mice. Expected orthologs of KIAA1340 include *R. norvegicus*, LOC312860 (86.9%), and *M. musculus*, C230080I20Rik (82.4%)

Unigene comparison of the KIAA1340 gene sequence indicates that the protein product of KIAA1340 demonstrates 33% identity, over a 155-aa length of aligned region, to mouse Kelch-like ECH-associated protein 1 (Cytosolic inhibitor of Nrf2); shares 32% identity, over a 155-amino-acid alignment, with rat Kelch-like ECH-associated protein 1 (Cytosolic inhibitor of Nrf2, INrf2); has 36% identity, over

98 aa, with the *A. thaliana* hypothetical protein F7A7.180; has 25% identity, over a 278-aa alignment, with *D. melanogaster* long-form kelch protein; and is 27% identical (over a 151-aa alignment) to *C. elegans* hypothetical protein W02G9.2.

Pfam analysis using a 410 amino acid fragment (SEQ ID NO: 44) of the predicted 505 amino acid KIAA1340 protein (SwissProt designation Q8N334) indicates KIAA1340 exhibits similarity to Kelch-related proteins and contains Kelch domains from amino acids 147-193 and 195-236, and partial Kelch domains from residues 238-277 and 291-340. Comparison domain mapping using the SMART database predicts Kelch domains from 159-206, 207-249, and 250-296. Identification of Kelch domains using SMART analysis reveals the Kelch domains to be located from residues 159-206, 207-249 and 250-296. The Kelch motif is an approximately 50-residue domain, named after the Drosophila mutant in which it was first identified, and is found six times in Drosophila egg-chamber regulatory protein, in mouse protein MIPP, in a number of poxviruses, in α- and β-scruin, and in galactose oxidase from the fungus *Dactylium dendroides*. The known functions of kelch-containing proteins include: scruin functioning as an actin cross-linking protein; galactose oxidase catalyzing the oxidation of the hydroxyl group at the C6 position in D-galactose; neuraminidase hydrolyzing sialic acid residues from glycoproteins; and kelch from Drosophila having a cytoskeletal function in the developing oocyte.

KIAA1340 also demonstrates a BTB/POZ domain at approximately amino acids 6-66 of SEQ ID NO: 44. The BTB (for BR-C, ttk and bab) or POZ (for Pox virus and Zinc finger) domain is present near the N-terminus of a fraction of zinc-finger proteins and in proteins that contain the motif, such as Kelch and a family of pox virus proteins. The BTB/POZ domain mediates homomeric dimerization and, in some instances, heteromeric dimerization. These domains are known to be protein-protein interaction domains found at the N-termini of several C2H2-type transcription factors, as well as Shaw-type potassium channels. POZ domains from several zinc-finger proteins have been shown to mediate transcriptional repression and to interact with components of histone deacetylase co-repressor complexes, including N-CoR and SMRT.

The presence of the Kelch domains in KIAA1340 indicates the protein may have several different functions, while the presence of the Zinc-binding domain

implicates KIAA1340 in protein-protein interactions, which are expected to be involved in the regulation of DNA binding and transcription.

FLJ20051: FLJ20051 exhibits a coding region of 2501 bp (Genbank Accession No. NM_019087; SEQ ID NOS: 31 and 32) and is expressed as a 4.3 kb transcript. Multiple Tissue Northern Blot (Clontech) of adult human RNA demonstrates that the FLJ20051 transcript is expressed at highest levels in kidney, placenta, heart, and skeletal muscle, and lower levels in the brain, lung, liver, colon, small intestine, thymus, peripheral blood leukocytes, and spleen.

Whole mount *in situ* analysis, as well as non-radioactive in situ hybridization (NR-ISH), of mouse mesenterium (E17.5, E18.5 or P6) detected FLJ20051 expression in lymphatic vessels of the gut mesenterium but not in arteries or veins, i.e., blood vessels. The antisense transcript, as detected by NR-ISH using the sense probe, is also expressed in the mesenteric lymphatic vessels, but to a lesser extent. These data indicate that the expression of FLJ20051 is controlled, at least in part, by its antisense transcript. These data also indicate that FLJ20051 plays a role in the development of the lymphatic endothelium.

Using *in situ* hybridization of frozen sections from E16.5 mouse embryos, FLJ20051 expression was detected in lung, kidney, brain and hair follicles in the skin. Multiple Tissue Northern Blot (Clontech) of adult mouse RNA revealed that the gene is expressed in kidney, and to a lesser extent in heart, brain and liver, and in still lesser extent in lung and testis, with no detectable expression in spleen or skeletal muscle.

Unigene comparison of the FLJ20051 gene sequence indicates that the predicted protein demonstrates 35% identity (over a 159-amino-acid alignment region) to a mouse homolog of human ADP-ribosylation factor 6; demonstrates 35% identity (over 159 aa) to the rat homolog of ADP-ribosylation factor 6; is 33% identical (over a 160-aa sequence alignment) to *A. thaliana* probable ADP-ribosylation factor At2g18390; shows 35% identity (over a 154-aa alignment) to ADP-ribosylation factor 3 of *D. melanogaster*; and is 34% identical (over a 168-aa sequence) to *C. elegans* ADP-ribosylation factors of the Arf family, which contains an ATP/GTP-binding P-loop. FLJ20051 orthologs are *R. norvegicus* LOC310106, (86.4% similarity), and *M. musculus* C230032K13Rik, (85.1% similarity).

Homology/Unigene data indicates that FLJ20051 acts as an ADP ribosylase. Domain mapping revealed that the 204-amino-acid protein (SwissProt designation Q9NXUS) contains GTP binding site from approximately residues 8-198. Pfam predicts an ADP-ribosylation factor (arf) domain from approximately residues 17 to 194 while SMART analysis predicts the GTPase region lies form amino acids 8-198. Small ADP ribosylation factor GTP-binding proteins are key regulators of vesicle biogenesis in intracellular trafficking. Members of the Arf family include Arl (Arf-like), Arp(Arf-related proteins) and the related Sar (Secretion-associated and Ras-related) proteins. Arf proteins cycle between inactive GDP-bound and active GTP-bound forms that bind selectively to effectors. The Arf domains can form complexes with GMP-PDE delta domains, which are believed to be specific soluble transport factors for certain prenylated proteins; Arl2-GTP is a regulator of PDE-mediated transport. Arf6 is found to be involved in cell motility, membrane dynamics, actin reorganization and tumor cell invasion.

Based on the foregoing information, FLJ20051 is expected to be an ADP Ribosylase active in lymphatic endothelial cells and to regulate GDP/GTP binding and intracellular trafficking, membrane dynamics, cell motility, or actin reorganization. FLJ20051 has recently been named ARFRP2: ADP-ribosylation factor related protein 2 (NM_019087).

FLJ13910: Microarray analysis of LEC-specific genes identified hypothetical protein FLJ13910 as having greater expression levels in LEC versus BEC. This gene was also detected in a putative LEC library. Further, the gene is expressed in lymphatic endothelial cells (LECs), but not in blood vascular endothelial cells (BECs) in an *in vitro* cell culture of primary endothelial cells (HDMEC, PromoCell), confirming that FLJ13910 is a LEC-specific gene.

Protein FLJ13910 (Genbank Accession No. NM_022780; SEQ ID NOS: 17 and 18) is encoded by an mRNA sequence of 3239 base pairs. Unigene analysis of the amino acid sequence of this mRNA indicates that FLJ13910 shows identity to the mouse erythroblast macrophage protein (26% identity over a 261-aa sequence), shows 37% identity (over 320 aa) to a putative *A. thaliana* WD repeat protein; and demonstrates 25% identity (within a 340-aa sequence) to *C. elegans*

T07D1.2.p. Species orthologs of FLJ13910 are found in *R. norvegicus* LOC312439 (94.5 %), and in *M. musculus* 1110007A06Rik (83.8 %).

Protein domain analysis of the 391-residue FLJ13910 protein (SwissProt designation Q9H871) revealed a LisH domain from amino acids 114-146 and a CTLH domain from approximately amino acids 153-210. The lissencephaly type-1-like homology motif (LisH) is an alpha-helical motif present in Lis1, treacle, Nopp140, some katanin p60 subunits, muskelin, tonneau, LEUNIG and WD40 repeat-containing proteins. It is believed that LisH motifs contribute to the regulation of microtubule dynamics, either by mediating dimerization, or by binding either a cytoplasmic dynein heavy chain or microtubules. The CTLH domain forms an alpha-helix. This domain is found in the mouse and human erythroblast macrophage proteins, both mouse and rat muskelin proteins, in the *A thaliana* putative WD repeat protein. Muskelin is a mediator of cell-adhesive behavior and cytoskeletal organization (Adams *et al., EMBO J.* 17:4964-74, 1998), whereas WD-repeat proteins are a group of structurally related proteins that participate in a wide range of cellular functions, including transmembrane signaling, mRNA modification, vesicle formation, and vesicular trafficking. Protein domain mapping predicts an E3 ubiquitin ligase domain (KOG2817)), which is indicative of post-translational modification, protein turnover or chaperone function.

Based on the protein domain mapping analysis, which indicates structural similarity to several actin binding or cytoskeleton related proteins, FLJ13910 is expected to act as a modulator of LEC migration and localization. FLJ13910 is also expected to act as an internal regulator of vesicle formation and vesicle trafficking, based on structural similarity to WD domain-containing proteins. Given the similarity to the muskelin protein, FLJ13910 is also expected to participate in structural and adhesion activities in LECs and plays a role in development and migration of LECs.

In vivo, NR-ISH of frozen sections from E16.5 mouse embryos detected gene transcript in the trigeminal ganglion, brain, nasal cavities, intestine, lung, submandibular gland and kidney.

FLJ11029: Protein FLJ11029 (Genbank Accession No. NM_018304; SEQ ID NOS: 33 and 34) is encoded by a 2334 bp mRNA sequence. Unigene

analysis of the amino acid sequence indicates that FLJ11029, which is upregulated in LEC compared to BEC, is a protein that shows no sequence similarity to any gene disclosed in a public database. Database exploration for species orthologs indicates that FLJ11029 is similar to *R. norvegicus* - LOC287597 (81.8 %), and *M. musculus* B930067F20Rik (81.6 %).

Protein domain mapping of the predicted 360 amino acid FLJ11029 (SwissProt designation Q9NUZ7) indicates that the amino acid sequence contains a domain at approximately residues 1-150 that is similar to domains associated with the PmbA_TldD family of proteins. It has been suggested that the TldD and PmbA proteins modulate the activity of DNA gyrase, or that PmbA may be involved in secretion. Pfam mapping also indicates FLJ11029 amino acids 151-360 are associated with domains found in multiple species of bacteria, the Bac surface Antigen, and are related to the UPF0140 family of eukaryotic proteins. Additional analysis indicates that this protein contains one of the following domains: Zinc finger, C3HC4 type (RING finger), ZPR1 zinc-finger domain, or C5HC2 zinc finger, which is expected to confer a DNA-binding function.

Protein domain mapping leads to the expectation that protein FLJ11029 acts as a DNA-binding/DNA-regulating protein involved in LEC transcriptional regulation.

KIAA1102: KIAA1102 (XM_044461; SEQ ID NO: 35 and 36) is encoded by a 6743 bp mRNA sequence. The gene is expressed in lymphatic endothelial cells (LECs), but not in blood vascular endothelial cells (BECs) in an *in vitro* cell culture of primary endothelial cells (HDMEC, PromoCell).

KIAA1102 is expressed as a 6.5 kb transcript in most tissues and as a 8.5 kb transcript in skeletal muscle. Expression in human tissue, as assayed by Northern blot analysis using Multiple Tissue Northern Blot (Clontech), is highest in heart, brain, skeletal muscle, and lung, with intermediate levels of expression found in prostate, thyroid, spleen, and placenta, and lower levels in kidney, small intestine, tongue and trachea. Assessment of KIAA1102 homologs in other species identified orthologs in *B. taurus* Bt.10388 (87.1%), *R. norvegicus* LOC305332 (84.0 %), and *M. musculus* LOC269654 (81.6%).

Analysis of protein similarity using the Unigene database reveals that KIAA1102 is 28% identical (over 308 amino acids) to human LIM domain only 7 isoform; is 23% identical (over a sequence of 197 aa) to mouse sciellin; is 24% identical (over 293 aa) to *C. elegans* F28F5.3b.p; and is 19% identical (over a sequence of 1371 aa) to the *D. melanogaster* T13564 microtubule-associated protein homolog.

Protein domain mapping of a 1101 fragment (SEQ ID NO. 45) of the predicted 1427 residue KIAA1102 protein (SwissProt designation Q9UPQ0) reveals similarity to a Calponin-like actin-binding domain (CH domain) from approximately residues 41-156; more particularly, Pfam analysis predicts the domain lies at residues 41-156 and SMART predicts a CH domain from amino acids 41-142. Mapping also detected a Zn-binding protein domain (LIM) spanning approximately amino acids 1030-1094 of SEQ ID NO: 45, and more specifically with Pfam analysis from aa 1031-1094, or with SMART analysis from residues 1030-1088.

The LIM domain is a conserved cysteine-rich domain of about 60 amino acids. This type of zinc-binding domain is present in diverse proteins, such as the homeobox protein Lim-1, the insulin gene enhancer binding protein Isl-1, and mammalian LH-2, a transcriptional regulatory protein involved in the control of cell differentiation in developing lymphoid and neural cell types. Some LIM domains bind protein partners via tyrosine-containing motifs. LIM domains are found in many key regulators of developmental pathways.

The calponin homology domain defines a superfamily of actin-binding domains found in both cytoskeletal proteins and signal transduction proteins, and is similar to a region in many actin-binding proteins, including spectrin, alpha-actinin and fimbrin, which contain a 250-amino-acid stretch called the actin binding domain(ABD). The ABD is believed to have arisen from duplication of a domain that is also found in as a single copy in a number of other proteins, like calponin or the vav proto-oncogene, and has been called the calponin homology (CH) domain. CH-domain-containing proteins are divided into monomeric actin cross-linking molecules, dimeric cross-linking proteins (e.g., alpha-actinin, beta-spectrin, and filamin) and monomeric F-actin binding proteins (dystrophin, utrophin). Each single ABD, comprising two CH domains, is able to bind one actin monomer. The CH domain

also occurs in a number of proteins not known to bind actin, a notable example being the vav-protooncogene. Examples of proteins containing CH domains include alpha-actinins, calponins, human dystrophin, mammalian vav proto-oncogene, human ras GTPase activating-like protein IQGAP1, Spectrin beta chain., Transgelins, and T-plastin.

Given the protein domain homology data above, it is expected that KIAA1102 functions as a DNA control protein using its Zinc binding domain and this protein may regulate development of LECs. Additionally, the KIAA1102 protein comprises a domain similar to those in actin-binding proteins and is expected to act as a regulator of cell motility.

In vivo, non-radioactive *in situ* hybridization (NR-ISH) of frozen sections from E16.5 mouse embryos showed KIAA1102 gene expression in lung, kidney glomerulus, thymus, brain, lips and bone marrow.*FLJ32029*: The protein-coding region of protein FLJ32029 (phosphoglucomutase 2-like 1 (PGM2L1), NM_173582; SEQ ID NOS: 37 and 38) lies within a 8456 bp mRNA sequence. The FLJ32029 mRNA is expressed as a 9.5 kb mRNA that is detected in brain and skeletal muscle. Comparison of the sequence to other mammalian species identified orthologs in *R. norvegicus* LOC293145 (84.9% similar) and *M. musculus* 4931406N15Rik (83.9%).

Unigene analysis revealed that FLJ32029 is 59% identical (over 596 aa) to human protein FLJ10983; 44% identical (over 600 aa) to a *C.elegans* phosphoglucomutase and phosphomannomutase phosphoserine; and 39% identical (over 576 aa) to *S. cerevisiae* protein YMR278w.

Mapping for potential protein family domain similarities indicates the predicted 622-amino-acid FLJ32029 protein (SwissProt designation Q96MQ7) comprises multiple phosphoglucomutase/phosphomannomutase domains (PGM-PMM-I and PGM-PMM II) at approximately amino acid residues 61-217 (alpha/beta/alpha domain I); amino acids 237-349 (alpha/beta/alpha domain II); amino acids 428-519 (alpha/beta/alpha domain III); and amino acids 520-617, similar to the phosphomannomutase C-terminal domain, which is involved in intramolecular transferase activity (e.g., phosphotransferase activity) and carbohydrate transport and metabolism. FLJ32029 is expected to function in carbohydrate regulation in

lymphatic endothelial cells and a dysfunctional FLJ32029 gene is expected to be involved in the development of lymphedema as a result of improper carbohydrate metabolism (Drubaix et al., *Int J Exp Pathol.* 78:117-21. 1997).

The identity of FLJ32029 as an LEC-specific gene was confirmed in an *in vitro* culture, where the gene was detected in lymphatic endothelial cells (LECs), and to a lesser extent in blood vascular endothelial cells (BECs) in cell culture of primary endothelial cells (HDMEC, PromoCell).

In vivo analysis of frozen sections from E16.5 mouse embryos (NR-ISH), detected gene expression in brain, bone marrow, glomeruli of the kidney cortex, thymus and intestine.*FLJ23594*: The protein-coding region of FLJ23594 (C18orf14, chromosome 18 open reading frame 14, NM_024781; SEQ ID NOS: 39 and 40) lies within a 2015 bp mRNA sequence. The 3 kb FLJ23594 transcript has been found to be expressed in human tissue at high levels in the placenta, lung, lymph node, and adrenal gland, as detected using a Multiple Tissue Northern Blot (Clontech) of adult human RNA. *In vitro* analysis shows that FLJ23594 was detected in lymphatic endothelial cells (LECs) and, to a lesser extent, in blood vascular endothelial cells (BECs) in cell culture of primary endothelial cells (HDMEC, PromoCell).

Comparison of the amino acid sequence with known amino acid sequences reveals that FLJ23594 shows 23% identity (over 271 aa) to human myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain, type B); demonstrates 22% identity (over 271 aa) to mouse myosin heavy chain, nonmuscle type B; is 22% identical (over a 271-aa sequence) to rat myosin heavy chain, nonmuscle type B; shows 24% identity (over 300 aa) to *D. melanogaster* myosin heavy chain 2; shows 23% identity (over 266 aa) to *S. cerevisiae* myosin-1 isoform (Type II myosin); shows 22% identity (over 269 aa) to an *A. thaliana* probable centromere protein; and is 22% identical (over a sequence of 284 aa) to *C. elegans* protein F52B10.1.

Protein domain mapping of FLJ23594 (SwissProt designation Q9H5C1) reveals a protein of 297 amino acids having two regions of similarity to bacterial M protein, from amino acids 148-168 and 216-236. Mapping also indicates regions of similarity, at aa 153-296, to the bZIP leucine zipper. The bZIP superfamily of eukaryotic DNA-binding transcription factors groups together proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper

required for dimerization. bZIP domains usually bind a pallindromic six-nucleotide site, but the specificity can be altered by interaction with an accessory factor. Some proteins that contain the bZIP domain include AP-1, oncogene v-jun, Jun-B and jun-D, probable transcription factors which are highly similar to jun/AP-1, and mammalian cAMP response element (CRE) binding proteins CREB, CREM, ATF-1, ATF-3, ATF-4, ATF-5, ATF-6 and LRF-1.

The similarity to bZIP-containing proteins indicates FLJ23594 acts as a DNA-binding protein and functions in the regulation of DNA transcription. The ability to regulate LEC-specific DNA-binding proteins or transcription factors allows for manipulation of LEC development, wherein the LEC can be induced to grow and proliferate by increased transcription or where the LEC can be induced to halt growth by blocking DNA regulatory functions. The structural similarity to cellular myosin also implicates FLJ23594 in cytokinesis, cell shaping, and specialized functions such as secretion and capping.

FLJ10532: Protein FLJ10532 (Genbank Accession No. AK001394; SEQ ID NO: 41) is encoded by a nucleic acid that was isolated as a 2372 bp mRNA fragment from fis, clone NT2RP2001044. The 2.3 kb transcript may be part of the message encoding protein DKFZp762K222. Cross-species comparison identified possible orthologs for DKFZp762K222 as *R. norvegicus* LOC306459 (88.3%), and *M. musculus* 4933409N07Rik (87.7 %). Cell culture of primary endothelial cells (HDMEC, PromoCell) indicates the FLJ10532 transcript is expressed in lymphatic endothelial cells (LECs) and is undetectable blood vascular endothelial cells (BECs).

BIKE: Identified herein as a LEC-specific gene, the human BIKE gene (Genbank Accession No. NM_017593; SEQ ID NOS: 25 and 26), a homolog of mouse BMP-2 kinase, exhibits a 2763 bp nucleotide sequence expressed as a 7.5 kb transcript. Northern blot analysis indicates that BIKE is expressed at high levels in skeletal muscle, tongue, lymph node, colon, and placenta, and at lower levels in heart, thyroid and stomach. BIKE demonstrates 93% identity, over a region of 329 aligned amino acids, to mouse Bmp2-inducible kinase; 39% identity, over a 289-residue alignment, to rat cyclin G-associated kinase; 40% identity, over a 270-aa alignment, to *Arabidopsis thaliana* hypothetical protein T21L14.21; 65% identity, over a 289-aa

region, to *D. melanogaster* Nak-P1; and 41% identity, over a 291-residue alignment, to a *C. elegans* serine/threonine kinase.

Protein domain mapping by Pfam analysis of a 661 residue human polypeptide sequence (SEQ ID NO: 46) comprising the predicted 355 amino acid BIKE protein [SwissProt designation Q9NSY1 (human), Q91Z96 (Mouse, 1138 residue fragment)] indicates the protein contains a protein kinase domain located from approximately residues 51 to 314. SMART domain predicts this domain lies from approximately residues 51-317. Thus, the protein kinase domain lies at approximately amino acids 51-317. SMART analysis indicates the domain is a STYKc domain, as in KIAA0472, which are expected to be phosphotransferases that exhibit dual-specificity as Ser/Thr/Tyr kinases.

The BIKE gene product is the human homolog of mouse BMP-2-inducible kinase. Bone morphogenic proteins (BMPs) play a key role in skeletal development and patterning. Expression of the mouse gene is increased during BMP-2-induced differentiation and the gene product is a putative serine/threonine protein kinase containing a nuclear localization signal. Therefore, the protein encoded by the human homolog is expected to be a protein kinase with a putative regulatory role in osteoblast differentiation. The unexpected identification of the BIKE protein as a LEC-specific gene reveals that the BMP homolog is involved in development of the lymphatic system, as well as osteogenesis.

The manipulation of lymphatic endothelial-cell-specific molecules is expected to be applicable to treatments of LEC diseases disorders associated with tissue edemas. Without wishing to be bound by theory, manipulation of such molecules is expected to modulate endothelial cell-cell or cell-matrix protein interactions and/or to affect transendothelial transport, thereby altering the state of fluid transport across the lymphatic vessel wall. Further, such molecules provide targets for the delivery of therapeutic compounds, such as growth factors, mitogens, and the like, as well as cytostatic or cytotoxic agents known in the art. These therapeutic compounds are targeted to such cells by associating a therapeutic agent with, e.g., a binding partner (such as an antibody) of the LEC surface marker. The

transmembrane proteins identified herein also provide useful targets for modulating cell adhesion or extracellular events integral to lymphatic endothelial cell function.

EXAMPLE 3

EX-VIVO CELL STIMULATION AND GENE THERAPY FOR LYMPHEDEMA WITH LEC-SPECIFIC GENE-TRANSFECTED CELLS

The ability of LEC-specific transcription factors to regulate genes specifically involved in LEC development provides a basis for treatment of individuals exhibiting a LEC disorder or condition resulting from either an increase or decrease in LEC gene expression levels. For example, mouse pups with a homozygous mutation in transcription factor Sox18 that disrupts formation of the Sox-18/MEF2C complex (Hosking, *et al., Biochem. Biophys. Res. Commun. 287*:493-500. 2001) develop chylous ascites in some genetic backgrounds (Pennisi, et al., *Nat. Genet. 24*:434-437. 2000), suggesting that both proteins may be involved in the regulation of lymphatic development. Additionally, upregulation of VEGFR-3 expression by transcription factor Prox-1 has been shown to be involved in development of LEC (See PCT/US03/06900 and Petrova et al., *EMBO J.* 21:4593-4599, 2002).

Similar to the effects of Prox-1, it is contemplated that upregulation of the LEC-specific genes disclosed herein is useful in promoting LEC development as a treatment for LEC disorders characterized by an under-developed lymphatic system. Conversely, decreasing activity of LEC-specific genes and downregulating LEC development is contemplated as a treatment for LEC disorders characterized by an over-developed lymphatic system such as lymphedema. Additionally, use of the genes or gene products disclosed herein as prophylactic therapeutics designed to serve as a therapeutic antagonist in therapies involving other primary agents is also contemplated, such as the use of a material disclosed herein to counter the influence of the primary agent on LEC gene expression. It is known in the art that *ex vivo* transfection of cells and subsequent transfer of these cells to patients is an effective method to upregulate *in vivo* levels of the transferred gene and to provide relief from a

disease resulting from under-expression of the gene(s) (Gelse et al., *Arthritis Rheum.* 48:430-41. 2003; Huard *et al, Gene Ther.* 9:1617-26. 2002; Kim *et al., Mol. Ther.* 6:591-600 2002).

To develop a therapy for treating irregularities of LEC development, endothelial cells, such as CAECs, SAVECs, LECs or BECs, are isolated from individuals experiencing a LEC disorder (e.g. lymphedema) and then placed in an appropriate culture medium (see above) to promote the growth and viability of the cells. The cells are then transfected with a viral vector designed to express any one of the LEC-specific genes to initiate LEC differentiation of the non-LECs *in vitro* and to promote growth of the LECs in culture. Transfection of cells with a viral vector, such as a replication-deficient adenovirus, adeno-associated virus, lentivurus or retrovirus, is well known to those of skill in the art (see, e.g., Laitinen *et al., Hum. Gene Ther.* 9:1481-1486, 1998; Petrova *et al., supra*)and is described above.

Cells transfected with a LEC-specific gene are then transferred into an affected patient in therapeutically effective numbers to promote LEC expansion *in vivo*. In preferred embodiments, the manipulated cells are autologous cells. These cells are delivered by one or more administrations typically involving injection. The cells are delivered at a local site of an LEC disease or disorder such as lymphedema; alternatively, the cells are delivered systemically, e.g., by injection remote to the perceived site in need of such cells or their product(s).

Administration of transfected cells to patients with lymphedema provides supplementary LECs that are incorporated into the lymphatic network, thereby promoting lymphatic development and lymph clearance to relieve the symptoms of lymphedema. It is contemplated that a method comprising LEC-specific gene transfection into endothelial cells and administration of transfected cells is useful in the treatment of any disease characterized by an alteration in LEC numbers or activity, such as lymphedema, lymphangioma, lymphangiomyeloma, lymphangiomatosis, lymphangiectasis, lymphosarcoma, and lymphangiosclerosis. Additionally, such methods are useful in ameliorating a symptom (e.g., lymph-induced swelling in the case of lymphedema) associated with such diseases.

EXAMPLE 4
MICROARRAY ANALYSIS TO DETECT LEC-AND LYMPH-RELATED DISORDERS

All LEC-specific genes identified, specifically any gene found in any of Tables 1 and 2 herein, are contemplated to be useful in the detection of LEC *in vivo* and in determining the extent of the lymphatic vasculature in a sample. The LEC-specific genes are also expected to be useful in diagnosing lymphedema and other LEC-related disorders. Further, the invention comprehends the use of a plurality of LEC-specific genes or their gene products for any of the applications identified herein. In this latter context, it is further contemplated that one or more LEC-specific genes, or their gene products, may be used in combination (regardless of whether such materials are physically or temporally combined or not) with at least one of the LEC-specific genes, or gene products, identified in Tables 3 and/or 4 of PCT/US03/06900.

Another aspect of the invention is a composition comprising a plurality of polynucleotide probes for use in detecting gene expression pattern(s) characteristic of particular cell type(s) and for detecting changes in the expression pattern of a particular cell type, *e.g.*, lymphatic endothelial cells. The term "polynucleotide probe" is used herein to refer to any one of the nucleic acid sequences listed in SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or any fragment thereof or a nucleic acid having a sequence encoding an amino acid sequence set forth in SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46, or a fragment thereof. Preferably, the fragment is at least 10 nucleotides in length; more preferably, it is at least 20 nucleotides in length. Such a composition is employed for the diagnosis and/or treatment of any condition or disease in which the dysfunction or non-function of lymphatic endothelial cells is implicated or suspected. The composition may alternatively include any one of genes found in Tables 3 and 4 of PCT/US03/06900 (specifically incorporated herein by reference). In one embodiment, the present invention provides a composition comprising a plurality of polynucleotide probes, wherein at least a subset of the polynucleotide probes comprises at least a portion of an expressed gene isolated from a population of LEC-specific genes identified above. Also contemplated is a composition comprising a plurality of polynucleotide probes, with

at least a subset of such probes each comprising a unique sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, and those sequences identified in any one of Tables 3 and 4 of PCT/US03/06900. Preferably, the composition comprises a subset of at least 3 polynucleotides, each having a different sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, and those sequences identified in any one of Tables 3 and 4 of PCT/US03/06900. Also preferred are compositions comprising at least 5, at least 7, at least 9, at least 15, at least 20, or at least 25 distinct polynucleotides having distinct sequences selected from the group of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, and those identified in any one of Tables 3 and 4 of PCT/US03/06900.

The composition is particularly useful as a set of hybridizable array elements in a microarray for monitoring the expression of a plurality of target polynucleotides. The microarray comprises a substrate and the hybridizable array elements. The microarray is used, for example, in the diagnosis and/or prognosis of a disease derived from aberrant lymphatic endothelial cell activity, such as lymphedema, lymphangioma, lymphangiomyeloma, lymphangiomatosis, lymphangiectasis, lymphosarcoma, and lymphangiosclerosis. Compositions may be useful in identifying more than one cell type and may be useful in the diagnosis and/or prognosis of more than one disease, disorder or condition. Further, useful information is obtained from those probes yielding a signal and from those probes not yielding a signal.

A polynucleotide comprising the sequence of any one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, and any one of the genes found in any of Tables 3 and 4 of PCT/US03/06900 may be used for the diagnosis of a condition, disorder, or disease in which the abnormal expression of any one of the genes encoded by SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, and any gene found in any one of Tables 3 and 4 of PCT/US03/06900, is associated. For example, a polynucleotide comprising any one of the sequences set forth in SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, and a gene found in any of Tables 3 and 4 of

PCT/US03/06900 may be used in hybridization or PCR assays of a fluid or tissue (e.g., obtained from a biopsy) to detect abnormal gene expression in patients with lymphedema or another lymph-associated disease, or suspected of having such a disease. In addition, a polynucleotide comprising a sequence encoding any of the amino acid sequences set forth in SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46, used in conjunction with at least one gene found in any of Tables 3 and 4 of PCT/US03/06900 is useful for the diagnosis of conditions or diseases associated with aberrant expression of a polypeptide having any one of those amino acid sequences. Fragments comprising at least 10 nucleotides are also useful in these diagnostic methods.

Expression profiles are generated using the compositions of the invention comprising, e.g., at least one sequence as set forth in SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 or 41. The expression profile generated from a microarray is used to detect changes in the expression of genes implicated in disease.

EXAMPLE 5

COMPOSITIONS OF THE INVENTION USEFUL FOR MODULATING VEGF-C/VEGFR-3 INTERACTIONS AND LEC DEVELOPMENT

VEGF-C and VEGF-D knockout mice demonstrate aberrant vascular development that can be overcome by administration of exogenous VEGF-C and/or VEGF-D polypeptide. To determine if transcriptional regulation by any one of the proteins having an amino acid sequence set forth in SEQ ID NOS: 18, 26, 28, 30, 32, 34, 36, 38 or 40 (which may act as LEC-specific transcription factors) can overcome this vascular defect due to its potential interaction with, and transcriptional effect on, the VEGFR-3 promoter, VEGF-C or VEGF-D knockout mice are genetically crossed by interbreeding with mice overexpressing a polypeptide having an amino acid sequence of any one of SEQ ID NOS: 18, 26, 28, 30, 32, 34, 36, 38, 40, 44, 45 or 46 from a cell-specific-promoter (e.g., a K-14 keratin promoter). The effects of transcription factor (TF) activity on lymphedema are assessed through measurement of lymphedema and vascular development, as described in Karkkainen et al. (*Proc. Natl. Acad. Sci. USA.* 98:12677-12682. 2001).

Survival of the knockout mice and detection of lymphatic development in the VEGF-C and/or VEGF-D knockout/TF-overexpressing mice indicates that the transcription factor of interest induces VEGFR-3 signaling and plays a role in lymphangiogenesis.

VEGF-C overexpressing mice (K-14-VEGF-C Tg) exhibit an extensive network of lymphatic vasculature, are prone to tumor metastasis, and demonstrate upregulated VEGFR-3 expression and symptoms of lymphedema (US Patent No. 6,361,946). To determine if the LEC-specific genes identified herein aid in regulating VEGF-C signaling through VEGFR-3, K-14-VEGF-C Tg mice are administered a composition of the invention, comprising at least one LEC-specific polypeptide or polypeptide fragment. Particularly useful are those fragments from extracellular proteins having sequences set forth in SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43.

Decreased lymphangiogenesis, decreased incidence of tumor metastasis, and decreased levels of VEGFR-3 exhibited by the LEC-specific-composition-treated K-14-VEGF-C mutant animals, as compared to the K-14-VEGF-C Tg single mutant animal, indicates that the administered composition interferes with VEGF-C signaling through VEGFR-3.

A result indicating that the composition of the invention is a negative regulator of lymphangiogenesis identifies a composition useful in a method of treating disorders mediated by extensive lymphatic vasculature, such as lymphangiogenesis in tumor development or lymphangiosarcoma. Such a composition, comprising at least one LEC-specific polypeptide, may be administered to a patient or an animal in need, thereby inhibiting or preventing the induction of lymphangiogenesis by interfering with lymphangiogenic signaling. A result indicating that a composition of the invention is a positive regulator of lymphatic endothelial cells identifies a composition suitable for use in a method of treating disorders characterized by insufficient LEC development, e.g., insufficient development of a lymphatic network.

Additionally, to determine the effects of the invention *in vivo*, compositions of the invention are administered to animals affected by lymphedema in the *Chy* mouse model of lymphedema (Lyon *et al., Mouse News Lett. 71*: 26, 1984). Lymphedema symptoms, such as footpad swelling and fluid retention/ascites, are

measured after treatment with either a soluble LEC-specific polypeptide, as described above, or with a LEC-specific gene therapy vector designed to regulate either cell surface proteins or intracellular proteins, comprising at least one of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 or 41, as disclosed previously. Improvement of lymphedema in the *Chy* mouse model after administration of compositions of the invention indicate that a composition which affects LEC-specific genes is capable of regulating LEC disorders, and is expected to be a useful therapeutic in the treatment of human lymphatic disorders.

Additionally contemplated is the treatment of subjects affected by lymphedema with a combination of therapeutics that includes a composition of the invention in conjunction with a lymphatic growth factor such as VEGF-C and/or VEGF-D to overcome impaired lymphatic vascular development.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

All of the above-cited U.S. patents, U.S. patent application publications, U.S. patent applications, foreign patents, foreign patent applications and non-patent publications referred to in this specification are incorporated herein by reference in their entireties.

CLAIMS

We claim:

1. A method for differentially modulating the growth or differentiation of blood endothelial cells (BEC) or lymphatic endothelial cells (LEC), comprising contacting endothelial cells with a composition comprising an agent that differentially modulates blood or lymphatic endothelial cells, said agent selected from the group consisting of:

(a) a polypeptide that comprises an amino acid sequence of a BEC polypeptide or a LEC polypeptide, or an active fragment of said polypeptide ;

(b) a polynucleotide that comprises a nucleotide sequence that encodes a polypeptide according to (a);

(c) an antibody that specifically binds to a polypeptide according to (a);

(d) a polypeptide comprising a fragment of the antibody of (c), wherein the fragment and the antibody bind to the polypeptide;

(e) an antisense nucleic acid to a human gene or mRNA encoding the polypeptide of (a);

(f) an interfering RNA (RNAi) to a human gene or mRNA encoding the polypeptide of (a).

2. A method according to claim 1, wherein the endothelial cells are contacted with the composition *ex vivo*.

3. A method according to claim 1, wherein the composition comprises a pharmaceutically acceptable diluent, adjuvant, or carrier, and the contacting step comprises administering the composition to a mammalian subject to differentially modulate BECs or LECs in the mammalian subject.

4. A method according to claim 3, comprising:

identifying a human subject with a disorder characterized by hyperproliferation of LECs; and

administering to the human subject the composition, wherein the agent differentially inhibits LEC growth compared to BEC growth.

5. A method according to claim 3, comprising:

identifying a human subject with a disorder characterized by hyperproliferation of LECs;

screening LECs of the subject to identify overexpression of a polypeptide set forth in any of Tables 1 and 2 herein; and

administering to the human subject the composition, wherein the agent differentially inhibits LEC growth compared to BEC growth by inhibiting expression of the polypeptide identified by the screening step.

6. A method according to claim 3 of modulating the growth of lymphatic endothelial cells in a human subject, comprising steps of:

identifying a human subject with a hypoproliferative lymphatic disorder;

screening the subject to identify underexpression or underactivity of a LEC polypeptide set forth in any of Tables 1 and 2 herein;

administering to the human subject said composition, wherein the agent comprises the LEC polypeptide (a) identified by the screening step or an active fragment of said polypeptide, or comprises the polynucleotide (b) that comprises a nucleotide sequence that encodes the polypeptide.

7. Use of an agent for the manufacture of a medicament for the differential modulation of blood vessel endothelial cell (BEC) or lymphatic vessel endothelial cell (LEC) growth or differentiation, said agent selected from the group consisting of:

(a) a polypeptide that comprises an amino acid sequence of a BEC polypeptide or a LEC polypeptide, or an active fragment of said polypeptide ;

(b) a polynucleotide that comprises a nucleotide sequence that encodes a polypeptide according to (a);

(c) an antibody that specifically binds to a polypeptide according to (a);

(d) a polypeptide comprising a fragment of the antibody of (c), wherein the fragment and the antibody bind to the polypeptide;

(e) an antisense nucleic acid to a human gene or mRNA encoding the polypeptide of (a);

(f) an interfering RNA (RNAi) to a human gene or mRNA encoding the polypeptide of (a).

8. A method or use according to any one of claims 1-7, wherein the polypeptide is a LEC polypeptide selected from the LEC polypeptides set forth in any of Tables 1 and 2 herein, and the agent differentially modulates LEC growth or differentiation over BEC growth or differentiation..

9. A method or use according to any one of claims 1-7, wherein the polypeptide is a BEC polypeptide selected from the BEC polypeptides set forth in Tables 1 and 2 herein, and the agent differentially modulates BEC growth or differentiation over LEC growth or differentiation.

10. The method according to any of claims 5 or 6, further comprising screening LECs of the subject to identify overexpression of a second polypeptide set forth in any of Tables 3 and 4 of PCT/US03/06900.

11. A method or use according to claim 8, wherein the LEC polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43.

12. A method or use according to claim 8, wherein the LEC polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 20.

13. A method or use according to claim 7, wherein the agent comprises an antibody according to (c) or polypeptide according to (d).

14. A method according to claim 7, wherein the agent comprises an extracellular domain fragment of the polypeptide of (a), or a polynucleotide encoding said extracellular domain fragment.

15. A method or use according to any one of claims 1-10, wherein the agent comprises an antisense molecule.

16. A method of treating hereditary lymphedema comprising:

identifying a human subject with lymphedema and with a mutation in at least one allele of a gene encoding a LEC protein identified in any of Tables 1 and 2, wherein the mutation correlates with lymphedema in human subjects, and with the proviso that said LEC protein is not VEGFR-3; and

administering to said subject a composition comprising a lymphatic growth agent selected from the group consisting of VEGF-C polypeptides, VEGF-D polypeptides, VEGF-C polynucleotides, and VEGF-D polynucleotides.

17. Use of a lymphatic growth agent selected from the group consisting of VEGF-C polypeptides, VEGF-D polypeptides, VEGF-C polynucleotides, and VEGF-D polynucleotides in the manufacture of a medicament for the treatment of hereditary lymphedema resulting from a mutation in a LEC gene identified in any of Tables 1 and 2.

18. A method of screening for an endothelial cell disorder or predisposition to said disorder, comprising:

obtaining a biological sample containing endothelial cell mRNA from a human subject; and

measuring expression of a BEC or LEC gene from the amount of mRNA in the sample transcribed from said gene, wherein the BEC or LEC gene encodes a polypeptide identified in any of Tables 1 and 2.

19. A method of monitoring the efficacy or toxicity of a drug on endothelial cells, comprising the steps of:

measuring expression of at least one BEC or LEC gene in endothelial cells of a mammalian subject before and after administering a drug to the subject, wherein the at least one BEC or LEC gene encodes a polypeptide set forth in any of Tables 1 and 2, and wherein changes in expression of the BEC or LEC gene correlates with efficacy or toxicity of the drug on endothelial cells.

20. A method of identifying compounds that modulate growth of endothelial cells, comprising

culturing endothelial cells in the presence and absence of a compound; and

measuring expression of at least one BEC or LEC gene in the cells, wherein the BEC or LEC gene is selected from the genes encoding polypeptides set forth in any of Tables 1 and 2, wherein a change in expression of at least one BEC gene in the presence compared to the absence of the compound identifies the compound as a modulator of BEC growth, and wherein a change in expression of at

least one LEC gene in the presence compared to the absence of the compound identifies the compound as a modulator of LEC growth.

21. A method according to claim 20 of screening for a compound that selectively modulates BEC or LEC growth or differentiation,

wherein the measuring step comprises measuring expression of at least one BEC gene and at least one LEC gene in the cells, and

wherein the method comprises screening for a compound that selectively modulates BEC or LEC growth or differentiation by selecting a compound that differentially modulates expression of the at least one BEC gene compared to expression of the at least one LEC gene.

22. A composition comprising

an isolated polynucleotide that comprises a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46; and

a pharmaceutically acceptable diluent, carrier or adjuvant.

23. A composition according to claim 22, comprising a polynucleotide that comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a fragment thereof that encodes the polypeptide.

24. An expression vector comprising an expression control sequence operably linked to a polynucleotide that comprises a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46.

25. An expression vector according to claim 24 that is a replication-deficient adenoviral or adeno-associated viral vector containing the polynucleotide.

26. A composition comprising an expression vector according to claim 24 or claim 25 and a pharmaceutically acceptable diluent, carrier, or adjuvant.

27. A kit comprising the composition according to any one of claims 22, 23, or 26 packaged with a protocol for administering the composition to a mammalian subject to modulate the lymphatic system in said subject.

28. A host cell transformed or transfected with an expression vector according to claim 24.

29. A method for producing a LEC polypeptide comprising the steps of growing a host cell according to claim 28 under conditions in which the cell expresses the polypeptide encoded by the polynucleotide.

30. A purified and isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46.

31. A purified and isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43; and

(b) an extracellular domain fragment of at least 10 amino acids of an amino acid sequence of (a).

32. A purified and isolated, soluble polypeptide according to claim 31 comprising an extracellular domain fragment of an amino acid sequence selected from the group consisting of : SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43, wherein the polypeptide lacks any transmembrane domain.

33. A polypeptide according to claim 32 that lacks any intracellular domain.

34. A fusion protein comprising a polypeptide according to claim 32 or 33 fused to an immunoglobulin fragment comprising an immunoglobulin constant region.

35. A composition comprising a polypeptide or protein according to any one of claims 30-34 and a pharmaceutically acceptable diluent, carrier or adjuvant.

36. A kit comprising the composition according to claim 35 and a protocol for administering said pharmaceutical composition to a mammalian subject to modulate the lymphatic system in said subject.

37. An antibody that specifically binds to a polypeptide according to any one of claims 30-34.

38. An antibody according to claim 37 that is a humanized antibody, a chimerized antibody, a human antibody, or a Fab fragment thereof.

39. A protein comprising an antigen binding domain of an antibody that specifically binds to a polypeptide according to any one of claim 30-34, wherein said protein specifically binds to said polypeptide.

40. A method of identifying a LEC nucleic acid comprising:

(a) contacting a biological sample containing a candidate LEC nucleic acid with a polynucleotide comprising a fragment of at least 14 contiguous nucleotides of a sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41, or a complement thereof, under the following stringent hybridization conditions:

(i) hybridization at 42°C for 20 hours in a solution containing 50% formamide, 5xSSPE, 5x Denhardt's solution, 0.1% SDS and 0.1 mg/ml denatured salmon sperm DNA, and

(ii) washing for 30 minutes at 65°C in 1xSSC, 0.1% SDS; and

(b) detecting hybridization of said candidate LEC nucleic acid and said polynucleotide, thereby identifying a LEC nucleic acid.

41. A method of identifying a LEC protein comprising:

(a) contacting a biological sample containing a candidate LEC protein with a LEC protein binding partner selected from the group consisting of an antibody according to claim 37 or a protein according to claim 39, under conditions suitable for binding therebetween; and

(b) detecting binding between said candidate LEC protein and said LEC binding partner, thereby identifying a LEC protein.

42. A method of identifying a LEC comprising:

(a) contacting a biological sample comprising cells with a LEC binding partner under conditions suitable for binding therebetween, wherein said LEC binding partner comprises an antibody that binds to a polypeptide comprising a sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43, or comprises an antigen binding fragment of said antibody; and

(b) identifying a LEC by detecting binding between a cell and said LEC binding partner, where binding of the LEC binding partner to the cell identifies a LEC.

43. A method of assaying for risk of developing hereditary lymphedema, comprising

(a) assaying nucleic acid of a human subject for a mutation that correlates with a hereditary lymphedema phenotype and alters the encoded amino acid sequence of at least one gene allele of the human subject when compared to the amino acid sequence of the polypeptide encoded by a corresponding wild-type gene allele, wherein the wild-type polypeptide is a polypeptide identified in any of Tables 1 and 2.

44. A method of assaying for risk of developing hereditary lymphedema, comprising

(a) assaying nucleic acid of a human subject for a mutation that correlates with a hereditary lymphedema phenotype and alters the encoded amino acid sequence of at least one gene allele of the human subject when compared to the amino acid sequence of the polypeptide encoded by a corresponding wild-type gene allele, wherein the wild-type polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43;

(b) correlating the presence or absence of said mutation in the nucleic acid to a risk of developing hereditary lymphedema, wherein the presence of said mutation in the nucleic acid correlates with an increased risk of developing hereditary lymphedema, and wherein the absence of said mutation in the nucleic acid correlates with no increased risk of developing hereditary lymphedema.

45. A method of assaying for risk of developing hereditary lymphedema, comprising

(a) assaying nucleic acid of a human subject for a mutation that alters the encoded amino acid sequence of at least one transcription factor allele of the human subject and alters transcription modulation activity of the transcription factor polypeptide encoded by the allele, when compared to the transcription modulation activity of a transcription factor polypeptide encoded by a wild-type allele,

wherein the wild-type transcription factor polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 18, 26, 28, 30, 32, 34, 36, 38, 40, 44, 45, and 46; and

(b) correlating the presence or absence of said mutation in the nucleic acid to a risk of developing hereditary lymphedema, wherein the presence of said mutation in the nucleic acid correlates with an increased risk of developing hereditary lymphedema, and wherein the absence of said mutation in the nucleic acid correlates with no increased risk of developing hereditary lymphedema.

46. The method according to claim 45 further comprising assaying said nucleic acid for a mutation that alters the expression of Sox18.

47. The method according to claim 46 wherein the assaying identifies a mutation altering activity of a transactivating or DNA binding domain of Sox18.

48. The method according to claim 46, wherein said mutation reduces transcriptional activation of a SOX18-responsive gene compared to transcriptional activation of said gene in the absence of said mutation.

49. A method of assaying for risk of developing hereditary lymphedema, comprising

(a) assaying nucleic acid of a human subject for a mutation that alters the encoded amino acid sequence of at least one LEC gene allele of the human subject and alters the binding affinity of the adhesion polypeptide encoded by the LEC gene allele, when compared to the binding affinity of an adhesion polypeptide encoded by a wild-type allele,

wherein the wild-type adhesion polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 20 and SEQ ID NO: 43; and

(b) correlating the presence or absence of said mutation in the nucleic acid to a risk of developing hereditary lymphedema, wherein the presence of said mutation in the nucleic acid correlates with an increased risk of developing hereditary lymphedema, and wherein the absence of said mutation in the nucleic acid correlates with no increased risk of developing hereditary lymphedema.

50. The method according to any one of claims 43-49, wherein the assaying identifies the presence of the mutation, and the correlating step identifies the increased risk of said patient developing hereditary lymphedema.

51. The method according to claim 49, further comprising assaying said nucleic acid for a mutation that alters the expression of Sox18.

52. A method of screening a human subject for an increased risk of developing hereditary lymphedema comprising assaying nucleic acid of a human subject for a mutation that alters the encoded amino acid sequence of at least one polypeptide comprising an amino acid sequence found in any of Tables 1 and 2.

53. The method according to claim 52, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 43, 44, 45, and 46 in a manner that correlates with the risk of developing hereditary lymphedema.

54. The method according to claim 52 further comprising assaying said nucleic acid for a mutation that alters the expression of Sox18.

55. The method according to any one of claims 43-54 wherein said method comprises at least one procedure selected from the group consisting of:

(a) determining a nucleotide sequence of at least one codon of at least one polynucleotide of the human subject;

(b) performing a hybridization assay to determine whether nucleic acid from the human subject has a nucleotide sequence identical to or different from one or more reference sequences;

(c) performing a polynucleotide migration assay to determine whether nucleic acid from the human subject has a nucleotide sequence identical to or different from one or more reference sequences; and

(d) performing a restriction endonuclease digestion to determine whether nucleic acid from the human subject has a nucleotide sequence identical to or different from one or more reference sequences.

56. The method according to any one of claims 43-54 wherein said method comprises: performing a polymerase chain reaction (PCR) to amplify nucleic acid comprising the coding sequence of said LEC polynucleotide, and determining nucleotide sequence of the amplified nucleic acid.

57. A method of screening for a hereditary lymphedema genotype in a human subject, comprising:

(a) providing a biological sample comprising nucleic acid from said subject, and

(b) analyzing said nucleic acid for the presence of a mutation altering the encoded amino acid sequence of the at least one allele of at least one gene in the human subject relative to a human gene encoding an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 40, 42, 43, 44, 45, and 46, wherein the presence of a mutation altering the encoded amino acid sequence in the human subject in a manner that correlates with lymphedema in human subjects identifies a hereditary lymphedema genotype.

58. The method according to claim 57 wherein said biological sample is a cell sample.

59. The method according to claim 57 wherein said analyzing comprises sequencing a portion of said nucleic acid.

60. The method according to claim 57 wherein the human subject has a hereditary lymphedema genotype identified by the method of screening.

61. A method of inhibiting lymphangiogenesis comprising administering to a subject an inhibitor of a LEC transmembrane polypeptide,

wherein the LEC transmembrane polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43; and

wherein the inhibitor is selected from the group consisting of

(a) a soluble extracellular domain fragment of the LEC transmembrane polypeptide;

(b) an antibody that binds to the extracellular domain of the LEC transmembrane polypeptide;

(c) a polypeptide comprising an antigen binding domain of the antibody according to (b); and

(d) an antisense nucleic acid complementary to the nucleic acid encoding the LEC transmembrane polypeptide or its complement.

62. A method according to claim 61, wherein the inhibitor is a polypeptide comprising an extracellular domain fragment of an LEC polypeptide, wherein the sequence of said extracellular domain is selected from the group consisting of amino acids 1-1005 of SEQ ID NO:43.

63. The method according to claim 61 wherein said subject is a human containing a tumor.

64. A method for modulating lymphangiogenesis in a mammalian subject comprising: administering to a mammalian subject in need of modulation of lymphangiogenesis an antisense molecule to a LEC polynucleotide, in an amount effective to inhibit transcription or translation of the polypeptide encoded by the LEC polynucleotide, wherein the LEC polynucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

65. A method of treating hereditary lymphedema, comprising:

(a) identifying a human subject with hereditary lymphedema and with a mutation that alters the encoded amino acid sequence of at least one polypeptide of the human subject, relative to the amino acid sequence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 6, 8, 10, 12, 14, 16, 20, 22, 24, 42 and 43; and

(b) administering to said subject a lymphatic growth factor selected from the group consisting of a VEGF-C polynucleotide, a VEGF-C polypeptide, a VEGF-D polynucleotide, and a VEGF-D polypeptide.

66. The method according to claim 8, further comprising a second LEC polypeptide selected from the LEC polypeptides set forth in any of Tables 3 and 4 of PCT/US03/06900.

67. The method according to any of claims 16 or 17, further comprising a mutation in at least one allele of a second gene encoding a LEC protein identified in any of Tables 3 and 4 of PCT/US03/06900, wherein the mutation in said second gene correlates with lymphedema in human subjects.

68. The method according to any of claims 18, 19, or 20, further comprising a second BEC or LEC gene that encodes a polypeptide set forth in any of Tables 3 and 4 of PCT/US03/06900.

69. The method according to any of claims 43 or 52, further comprising assaying the nucleic acid for a second mutation that correlates with a hereditary lymphedema phenotype and alters the encoded amino acid sequence of a second gene allele of the human subject when compared to the amino acid sequence of the polypeptide encoded by a corresponding wild-type allele of said second gene, wherein the wild-type polypeptide encoded by said second gene is a polypeptide identified in any of Tables 3 and 4 of PCT/US03/06900.

70. An isolated polypeptide comprising an amino acid sequence at least 95% identical to amino acids 1-1005 of SEQ ID NO: 43.

71. A soluble polypeptide comprising a fragment of the amino acid sequence set forth in SEQ ID NO: 43, wherein said fragment lacks the transmembrane and intracellular amino acids from approximately residues 1006-1028 of SEQ ID NO: 43.

72. An isolated polypeptide comprising an adhesion domain having an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 20 or SEQ ID NO: 43, wherein said fragment includes at least one vonWillebrand factor domain.

73. An isolated polypeptide comprising a Cache domain and comprising an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 20 or SEQ ID NO: 43, wherein said fragment includes at least one Cache domain.

74. A fusion protein comprising a polypeptide according to any one of claims 70-73 and a heterologous polypeptide.

75. An antibody that specifically binds to a polypeptide according to any one of claims 70-73.

76. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a protein kinase domain selected from the group consisting of a protein kinase domain set out in SEQ ID NO: 8, in SEQ ID NO: 26, at approximately amino acids 88-342 of SEQ ID NO: 42 and at approximately amino acids 51-317 of SEQ ID NO: 46.

77. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a Zinc binding domain selected from the group consisting of a Zinc-binding domain set out in SEQ ID NO: 36, a Zinc-binding domain at approximately amino acids 527-561 of SEQ ID NO: 6, and a Zinc-binding domain at approximately amino acids 1030-1094 of SEQ ID NO: 45.

78. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a transcription factor selected from the group consisting of SEQ ID NO: 30, SEQ ID NO: 38, SEQ ID NO: 40, and SEQ ID NO: 44.

79. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a cytoskeleton regulation protein selected from the group consisting of SEQ ID NO: 18, SEQ ID NO: 36, and SEQ ID NO: 45.

80. An isolated polypeptide comprising a SPRY domain having an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 6, wherein said fragment includes at least one SPRY domain.

81. An isolated polypeptide comprising a cation transport activity and comprising an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 12, wherein said fragment includes at least one cation transport domain.

82. An isolated polypeptide comprising a guanine-nucleotide dissociation stimulator activity and comprising an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 28, wherein said fragment includes at least one RCC1-like domain.

83. An isolated polypeptide comprising a ubitquitin ligase activity and comprising an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 28, wherein said fragment includes at least one Hect domain.

84. An isolated polypeptide comprising a Kelch domain and comprising an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 30 or SEQ ID NO: 44, wherein said fragment includes at least one Kelch domain

85. An isolated polypeptide comprising a GTP-binding activity and comprising an amino acid sequence at least 95% identical to a fragment of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 32, wherein said fragment includes at least one GTP-binding domain.

86. A polynucleotide comprising a nucleotide sequence that encodes a polypeptide according to any one of claims 70-85.

87. An expression vector comprising a polynucleotide according to claim 85 operatively linked to an expression control sequence.

88. An expression vector according to claim 87 that is a replication-deficient adenoviral vector or adeno-associated virus vector.

89. A composition comprising the polypeptide of any one of claims 70-87 for use in the methods of the invention.

SEQUENCE LISTING

<110> Alitalo, et al.

<120> Lymphatic and Blood Endothelial Cell Genes

<130> 28967/39230

<160> 46

<170> PatentIn version 3.2

<210> 1
<211> 1997
<212> DNA
<213> Homo sapiens

<400> 1
cccgccccgc ctctccaaaa agctacaccg acgcggaccg cggcggcgtc ctccctcgcc 60

ctcgcttcac ctcgcgggct ccgaatgcgg ggagctcgga tgtccggttt cctgtgaggc 120

ttttacctga cacccgccgc ctttcccccgg cactggctgg gagggcgccc tgcaaagttg 180

ggaacgcgga gccccggacc cgctcccgcc gcctccggct cgcccagggg gggtcgccgg 240

gaggagcccg ggggagaggg accaggaggg gcccgcggcc tcgcaggggc gcccgcgccc 300

ccacccctgc ccccgccagc ggaccggtcc cccacccccg gtccttccac catgcacttg 360

ctgggcttct tctctgtggc gtgttctctg ctcgccgctg cgctgctccc gggtcctcgc 420

gaggcgcccg ccgccgccgc cgccttcgag tccggactcg acctctcgga cgcggagccc 480

gacgcgggcg aggccacggc ttatgcaagc aaagatctgg aggagcagtt acggtctgtg 540

tccagtgtag atgaactcat gactgtactc tacccagaat attggaaaat gtacaagtgt 600

cagctaagga aaggaggctg gcaacataac agagaacagg ccaacctcaa ctcaaggaca 660

gaagagacta taaaatttgc tgcagcacat tataatacag agatcttgaa aagtattgat 720

aatgagtgga gaaagactca atgcatgcca cgggaggtgt gtatagatgt ggggaaggag 780

tttggagtcg cgacaaacac cttctttaaa cctccatgtg tgtccgtcta cagatgtggg 840

ggttgctgca atagtgaggg gctgcagtgc atgaacacca gcacgagcta cctcagcaag 900

acgttatttg aaattacagt gcctctctct caaggcccca aaccagtaac aatcagtttt 960

gccaatcaca cttcctgccg atgcatgtct aaactggatg tttacagaca agttcattcc 1020

attattagac gttccctgcc agcaacacta ccacagtgtc aggcagcgaa caagacctgc 1080

cccaccaatt acatgtggaa taatcacatc tgcagatgcc tggctcagga agattttatg 1140

ttttcctcgg atgctggaga tgactcaaca gatggattcc atgacatctg tggaccaaac 1200

aaggagctgg atgaagagac ctgtcagtgt gtctgcagag cggggcttcg gcctgccagc 1260

tgtggacccc acaaagaact agacagaaac tcatgccagt gtgtctgtaa aaacaaactc 1320

ttccccagcc aatgtggggc caaccgagaa tttgatgaaa acacatgcca gtgtgtatgt 1380

aaaagaacct gccccagaaa tcaaccccta aatcctggaa aatgtgcctg tgaatgtaca 1440

gaaagtccac agaaatgctt gttaaaagga aagaagttcc accaccaaac atgcagctgt 1500

tacagacggc catgtacgaa ccgccagaag gcttgtgagc caggattttc atatagtgaa 1560

gaagtgtgtc gttgtgtccc ttcatattgg aaaagaccac aaatgagcta agattgtact 1620

gttttccagt tcatcgattt tctattatgg aaaactgtgt tgccacagta gaactgtctg 1680

tgaacagaga gacccttgtg ggtccatgct aacaaagaca aaagtctgtc tttcctgaac 1740

catgtggata actttacaga aatggactgg agctcatctg caaaaggcct cttgtaaaga 1800

ctggttttct gccaatgacc aaacagccaa gattttcctc ttgtgatttc tttaaaagaa 1860

tgactatata atttatttcc actaaaaata ttgtttctgc attcattttt atagcaacaa 1920

caattggtaa aactcactgt gatcaatatt tttatatcat gcaaaatatg tttaaaataa 1980

aatgaaaatt gtattat 1997

<210> 2
<211> 419
<212> PRT
<213> Homo sapiens

<400> 2

Met His Leu Leu Gly Phe Phe Ser Val Ala Cys Ser Leu Leu Ala Ala
1 5 10 15

Ala Leu Leu Pro Gly Pro Arg Glu Ala Pro Ala Ala Ala Ala Ala Phe
20 25 30

Glu Ser Gly Leu Asp Leu Ser Asp Ala Glu Pro Asp Ala Gly Glu Ala
35 40 45

Thr Ala Tyr Ala Ser Lys Asp Leu Glu Glu Gln Leu Arg Ser Val Ser
50 55 60

Ser Val Asp Glu Leu Met Thr Val Leu Tyr Pro Glu Tyr Trp Lys Met
65 70 75 80

Tyr Lys Cys Gln Leu Arg Lys Gly Gly Trp Gln His Asn Arg Glu Gln
85 90 95

Ala Asn Leu Asn Ser Arg Thr Glu Glu Thr Ile Lys Phe Ala Ala Ala
100 105 110

His Tyr Asn Thr Glu Ile Leu Lys Ser Ile Asp Asn Glu Trp Arg Lys
115 120 125

Thr Gln Cys Met Pro Arg Glu Val Cys Ile Asp Val Gly Lys Glu Phe

130 135 140

Gly Val Ala Thr Asn Thr Phe Phe Lys Pro Pro Cys Val Ser Val Tyr
145 150 155 160

Arg Cys Gly Gly Cys Cys Asn Ser Glu Gly Leu Gln Cys Met Asn Thr
165 170 175

Ser Thr Ser Tyr Leu Ser Lys Thr Leu Phe Glu Ile Thr Val Pro Leu
180 185 190

Ser Gln Gly Pro Lys Pro Val Thr Ile Ser Phe Ala Asn His Thr Ser
195 200 205

Cys Arg Cys Met Ser Lys Leu Asp Val Tyr Arg Gln Val His Ser Ile
210 215 220

Ile Arg Arg Ser Leu Pro Ala Thr Leu Pro Gln Cys Gln Ala Ala Asn
225 230 235 240

Lys Thr Cys Pro Thr Asn Tyr Met Trp Asn Asn His Ile Cys Arg Cys
245 250 255

Leu Ala Gln Glu Asp Phe Met Phe Ser Ser Asp Ala Gly Asp Asp Ser
260 265 270

Thr Asp Gly Phe His Asp Ile Cys Gly Pro Asn Lys Glu Leu Asp Glu
275 280 285

Glu Thr Cys Gln Cys Val Cys Arg Ala Gly Leu Arg Pro Ala Ser Cys
290 295 300

Gly Pro His Lys Glu Leu Asp Arg Asn Ser Cys Gln Cys Val Cys Lys
305 310 315 320

Asn Lys Leu Phe Pro Ser Gln Cys Gly Ala Asn Arg Glu Phe Asp Glu
325 330 335

Asn Thr Cys Gln Cys Val Cys Lys Arg Thr Cys Pro Arg Asn Gln Pro
340 345 350

Leu Asn Pro Gly Lys Cys Ala Cys Glu Cys Thr Glu Ser Pro Gln Lys
355 360 365

Cys Leu Leu Lys Gly Lys Lys Phe His His Gln Thr Cys Ser Cys Tyr
370 375 380

Arg Arg Pro Cys Thr Asn Arg Gln Lys Ala Cys Glu Pro Gly Phe Ser
385 390 395 400

Tyr Ser Glu Glu Val Cys Arg Cys Val Pro Ser Tyr Trp Lys Arg Pro
405 410 415

Gln Met Ser

<210> 3
<211> 2781
<212> DNA
<213> Homo sapiens

<400> 3
atggatatgt ttcctctcac ctgggttttc ttagccctct acttttcaag acaccaagtg 60

agaggccaac cagacccacc gtgcggaggt cgtttgaatt ccaaagatgc tggctatatc 120

acctctcccg gttaccccca ggactacccc tcccaccaga actgcgagtg gattgtttac 180

gcccccgaac ccaaccagaa gattgtcctc aacttcaacc ctcactttga aatcgagaag 240

cacgactgca agtatgactt tatcgagatt cgggatgggg acagtgaatc cgcagacctc 300

ctgggcaaac actgtgggaa catcgccccg cccaccatca tctcctcggg ctccatgctc 360

tacatcaagt tcacctccga ctacgcccgg caggggcag gcttctctct gcgctacgag 420

atcttcaaga caggctctga agattgctca aaaaacttca caagccccaa cgggaccatc 480

gaatctcctg ggtttcctga gaagtatcca cacaacttgg actgcacctt taccatcctg 540

gccaaaccca agatggagat catcctgcag ttcctgatct ttgacctgga gcatgaccct 600

ttgcaggtgg gagaggggga ctgcaagtac gattggctgg acatctggga tggcattcca 660

catgttggcc ccctgattgg caagtactgt gggaccaaaa caccctctga acttcgttca 720

tcgacgggga tcctctccct gacctttcac acggacatgg cggtggccaa ggatggcttc 780

tctgcgcgtt actacctggt ccaccaagag ccactagaga actttcagtg caatgttcct 840

ctgggcatgg agtctggccg gattgctaat gaacagatca gtgcctcatc tacctactct 900

gatgggaggt ggacccctca acaaagccgg ctccatggtg atgacaatgg ctggaccccc 960

aacttggatt ccaacaagga gtatctccag gtggacctgc gctttttaac catgctcacg 1020

gccatcgcaa cacagggagc gatttccagg gaaacacaga atggctacta cgtcaaatcc 1080

tacaagctgg aagtcagcac taatggagag gactggatgg tgtaccggca tggcaaaaac 1140

cacaaggtat ttcaagccaa caacgatgca actgaggtgg ttctgaacaa gctccacgct 1200

ccactgctga caaggtttgt tagaatccgc cctcagacct ggcactcagg tatcgccctc 1260

cggctggagc tcttcggctg ccgggtcaca gatgctccct gctccaacat gctggggatg 1320

ctctcaggcc tcattgcaga ctcccagatc tccgcctctt ccacccagga atacctctgg 1380

agccccagtg cagcccgcct ggtcagcagc cgctcgggct ggttccctcg aatccctcag 1440

gcccagcccg gtgaggagtg gcttcaggta gatctgggaa cacccaagac agtgaaaggt 1500

gtcatcatcc agggagcccg cggaggagac agtatcactg ctgtggaagc cagagcattt 1560

gtgcgcaagt tcaaagtctc ctacagccta aacggcaagg actgggaata cattcaggac 1620

cccaggaccc agcagccaaa gctgttcgaa gggaacatgc actatgacac ccctgacatc 1680

cgaaggtttg accccattcc ggcacagtat gtgcgggtat acccggagag gtggtcgccg 1740

gcggggattg ggatgcggct ggaggtgctg ggctgtgact ggacagactc caagcccacg 1800

gtaaaaacgc tgggacccac tgtgaagagc gaagagacaa ccacccccta ccccaccgaa 1860

gaggaggcca cagagtgtgg ggagaactgc agctttgagg atgacaaaga tttgcagctc 1920

ccttcgggat tcaattgcaa cttcgatttc ctcgaggagc cctgtggttg gatgtatgac 1980

catgccaagt ggctccggac cacctgggcc agcagctcca gcccaaacga ccggacgttt 2040

ccagatgaca ggaatttctt gcggctgcag agtgacagcc agagagaggg ccagtatgcc 2100

cggctcatca gccccccctgt ccacctgccc cgaagcccgg tgtgcatgga gttccagtac 2160

caggccacgg gcggccgcgg ggtggcgctg caggtggtgc gggaagccag ccaggagagc 2220

aagttgctgt gggtcatccg tgaggaccag ggcggcgagt ggaagcacgg gcggatcatc 2280

ctgcccagct acgacatgga gtaccagatt gtgttcgagg gagtgatagg gaaaggacgt 2340

tccggagaga ttgccattga tgacattcgg ataagcactg atgtcccact ggagaactgc 2400

atggaaccca tctcggcttt tgcagtggac atcccagaaa tacatgagag agaaggatat 2460

gaagatgaaa ttgatgatga atacgaggtg gactggagca attcttcttc tgcaacctca 2520

gggtctggcg ccccctcgac cgacaaagaa aagagctggc tgtacaccct ggatcccatc 2580

ctcatcacca tcatcgccat gagctcactg ggcgtcctcc tgggggccac ctgtgcaggc 2640

ctcctgctct actgcacctg ttcctactcg ggcctgagct cccgaagctg caccacactg 2700

gagaactaca acttcgagct ctacgatggc cttaagcaca aggtcaagat gaaccaccaa 2760

aagtgctgct ccgaggcatg a 2781

<210> 4
<211> 926
<212> PRT
<213> Homo sapiens

<400> 4

Met Asp Met Phe Pro Leu Thr Trp Val Phe Leu Ala Leu Tyr Phe Ser
1 5 10 15

Arg His Gln Val Arg Gly Gln Pro Asp Pro Pro Cys Gly Gly Arg Leu
20 25 30

```
Asn Ser Lys Asp Ala Gly Tyr Ile Thr Ser Pro Gly Tyr Pro Gln Asp
        35                  40                  45

Tyr Pro Ser His Gln Asn Cys Glu Trp Ile Val Tyr Ala Pro Glu Pro
    50                  55                  60

Asn Gln Lys Ile Val Leu Asn Phe Asn Pro His Phe Glu Ile Glu Lys
65                  70                  75                  80

His Asp Cys Lys Tyr Asp Phe Ile Glu Ile Arg Asp Gly Asp Ser Glu
                85                  90                  95

Ser Ala Asp Leu Leu Gly Lys His Cys Gly Asn Ile Ala Pro Pro Thr
            100                 105                 110

Ile Ile Ser Ser Gly Ser Met Leu Tyr Ile Lys Phe Thr Ser Asp Tyr
        115                 120                 125

Ala Arg Gln Gly Ala Gly Phe Ser Leu Arg Tyr Glu Ile Phe Lys Thr
    130                 135                 140

Gly Ser Glu Asp Cys Ser Lys Asn Phe Thr Ser Pro Asn Gly Thr Ile
145                 150                 155                 160

Glu Ser Pro Gly Phe Pro Glu Lys Tyr Pro His Asn Leu Asp Cys Thr
                165                 170                 175

Phe Thr Ile Leu Ala Lys Pro Lys Met Glu Ile Ile Leu Gln Phe Leu
            180                 185                 190

Ile Phe Asp Leu Glu His Asp Pro Leu Gln Val Gly Glu Gly Asp Cys
        195                 200                 205

Lys Tyr Asp Trp Leu Asp Ile Trp Asp Gly Ile Pro His Val Gly Pro
    210                 215                 220

Leu Ile Gly Lys Tyr Cys Gly Thr Lys Thr Pro Ser Glu Leu Arg Ser
225                 230                 235                 240

Ser Thr Gly Ile Leu Ser Leu Thr Phe His Thr Asp Met Ala Val Ala
                245                 250                 255

Lys Asp Gly Phe Ser Ala Arg Tyr Tyr Leu Val His Gln Glu Pro Leu
            260                 265                 270

Glu Asn Phe Gln Cys Asn Val Pro Leu Gly Met Glu Ser Gly Arg Ile
        275                 280                 285
```

```
Ala Asn Glu Gln Ile Ser Ala Ser Ser Thr Tyr Ser Asp Gly Arg Trp
     290                 295                 300

Thr Pro Gln Gln Ser Arg Leu His Gly Asp Asp Asn Gly Trp Thr Pro
305                 310                 315                 320

Asn Leu Asp Ser Asn Lys Glu Tyr Leu Gln Val Asp Leu Arg Phe Leu
                325                 330                 335

Thr Met Leu Thr Ala Ile Ala Thr Gln Gly Ala Ile Ser Arg Glu Thr
            340                 345                 350

Gln Asn Gly Tyr Tyr Val Lys Ser Tyr Lys Leu Glu Val Ser Thr Asn
        355                 360                 365

Gly Glu Asp Trp Met Val Tyr Arg His Gly Lys Asn His Lys Val Phe
    370                 375                 380

Gln Ala Asn Asn Asp Ala Thr Glu Val Val Leu Asn Lys Leu His Ala
385                 390                 395                 400

Pro Leu Leu Thr Arg Phe Val Arg Ile Arg Pro Gln Thr Trp His Ser
                405                 410                 415

Gly Ile Ala Leu Arg Leu Glu Leu Phe Gly Cys Arg Val Thr Asp Ala
            420                 425                 430

Pro Cys Ser Asn Met Leu Gly Met Leu Ser Gly Leu Ile Ala Asp Ser
        435                 440                 445

Gln Ile Ser Ala Ser Ser Thr Gln Glu Tyr Leu Trp Ser Pro Ser Ala
    450                 455                 460

Ala Arg Leu Val Ser Ser Arg Ser Gly Trp Phe Pro Arg Ile Pro Gln
465                 470                 475                 480

Ala Gln Pro Gly Glu Glu Trp Leu Gln Val Asp Leu Gly Thr Pro Lys
                485                 490                 495

Thr Val Lys Gly Val Ile Ile Gln Gly Ala Arg Gly Gly Asp Ser Ile
            500                 505                 510

Thr Ala Val Glu Ala Arg Ala Phe Val Arg Lys Phe Lys Val Ser Tyr
        515                 520                 525

Ser Leu Asn Gly Lys Asp Trp Glu Tyr Ile Gln Asp Pro Arg Thr Gln
    530                 535                 540
```

```
Gln Pro Lys Leu Phe Glu Gly Asn Met His Tyr Asp Thr Pro Asp Ile
545                 550                 555                 560

Arg Arg Phe Asp Pro Ile Pro Ala Gln Tyr Val Arg Val Tyr Pro Glu
                565                 570                 575

Arg Trp Ser Pro Ala Gly Ile Gly Met Arg Leu Glu Val Leu Gly Cys
            580                 585                 590

Asp Trp Thr Asp Ser Lys Pro Thr Val Lys Thr Leu Gly Pro Thr Val
        595                 600                 605

Lys Ser Glu Glu Thr Thr Thr Pro Tyr Pro Thr Glu Glu Glu Ala Thr
    610                 615                 620

Glu Cys Gly Glu Asn Cys Ser Phe Glu Asp Asp Lys Asp Leu Gln Leu
625                 630                 635                 640

Pro Ser Gly Phe Asn Cys Asn Phe Asp Phe Leu Glu Glu Pro Cys Gly
                645                 650                 655

Trp Met Tyr Asp His Ala Lys Trp Leu Arg Thr Thr Trp Ala Ser Ser
            660                 665                 670

Ser Ser Pro Asn Asp Arg Thr Phe Pro Asp Asp Arg Asn Phe Leu Arg
        675                 680                 685

Leu Gln Ser Asp Ser Gln Arg Glu Gly Gln Tyr Ala Arg Leu Ile Ser
    690                 695                 700

Pro Pro Val His Leu Pro Arg Ser Pro Val Cys Met Glu Phe Gln Tyr
705                 710                 715                 720

Gln Ala Thr Gly Gly Arg Gly Val Ala Leu Gln Val Val Arg Glu Ala
                725                 730                 735

Ser Gln Glu Ser Lys Leu Leu Trp Val Ile Arg Glu Asp Gln Gly Gly
            740                 745                 750

Glu Trp Lys His Gly Arg Ile Ile Leu Pro Ser Tyr Asp Met Glu Tyr
        755                 760                 765

Gln Ile Val Phe Glu Gly Val Ile Gly Lys Gly Arg Ser Gly Glu Ile
    770                 775                 780

Ala Ile Asp Asp Ile Arg Ile Ser Thr Asp Val Pro Leu Glu Asn Cys
785                 790                 795                 800
```

Met Glu Pro Ile Ser Ala Phe Ala Val Asp Ile Pro Glu Ile His Glu
805 810 815

Arg Glu Gly Tyr Glu Asp Glu Ile Asp Asp Glu Tyr Glu Val Asp Trp
820 825 830

Ser Asn Ser Ser Ser Ala Thr Ser Gly Ser Gly Ala Pro Ser Thr Asp
835 840 845

Lys Glu Lys Ser Trp Leu Tyr Thr Leu Asp Pro Ile Leu Ile Thr Ile
850 855 860

Ile Ala Met Ser Ser Leu Gly Val Leu Leu Gly Ala Thr Cys Ala Gly
865 870 875 880

Leu Leu Leu Tyr Cys Thr Cys Ser Tyr Ser Gly Leu Ser Ser Arg Ser
885 890 895

Cys Thr Thr Leu Glu Asn Tyr Asn Phe Glu Leu Tyr Asp Gly Leu Lys
900 905 910

His Lys Val Lys Met Asn His Gln Lys Cys Cys Ser Glu Ala
915 920 925

<210> 5
<211> 3540
<212> DNA
<213> Homo sapiens

<400> 5
ggagggctgg ggcggtacga agcgggggtg ggctctgcgc gtaatggcag cgccgtggcc 60

tcgcgtccat ctttgccgtt ctctcggacc tgtcacaaag gagtcgcgcc gccgccgccg 120

ccccctccct ccggtgggcc cgggaggtag agaaaaactg ccattggatg tccagaatcc 180

cctgtagttg ataatgttgg gaataagctc tgcaactttc tttggcattc agttgttaaa 240

aacaaatagg atgcaaattc ctcaactcca ggttatgaaa acagtacttg gaaaactgaa 300

aactacctaa atgatcgtct ttggttgggc cgtgttctta gcgagcagaa gccttggcca 360

gggtctgttg ttgactctcg aagagcacat agcccacttc ctagggactg gaggtgccgc 420

tactaccatg ggtaattcct gtatctgccg agatgacagt ggaacagatg acagtgttga 480

cacccaacag caacaggccg agaacagtgc agtacccact gctgacacaa ggagccaacc 540

acgggaccct gttcggccac caaggagggg ccgaggacct catgagccaa ggagaaagaa 600

acaaaatgtg gatgggctag tgttggacac actggcagta atacggactc ttgtagataa 660

tgatcaggaa cctccctatt caatgataac attacacgaa atggcagaaa cagatgaagg 720

atggttggat gttgtccagt ctttaattag agttattcca ctggaagatc cactgggacc 780

agctgttata acattgttac tagatgaatg tccattgccc actaaagatg cactccagaa 840
attgactgaa attctcaatt taaatggaga agtagcttgc caggactcaa gccatcctgc 900
caaacacagg aacacatctg cagtcctagg ctgcttggcc gagaaactag caggtcctgc 960
aagtataggt ttacttagcc caggaatact ggaatacttg ctacagtgtc tgaagttaca 1020
gtcccacccc acagtcatgc tttttgcact tatcgcactg gaaaagtttg cacagacaag 1080
tgaaaataaa ttgactattt ctgaatccag tattagtgac cggcttgtca cattggagtc 1140
ctgggctaat gatcctgatt atctgaaacg tcaagttggt ttctgtgccc agtggagctt 1200
agacaatctc tttttaaaag aaggtagaca gctgacctat gagaaagtga acttgagtag 1260
cattagggcc atgctgaata gcaatgatgt cagcgagtac ctgaagatct cacctcatgg 1320
cttagaggct cgctgtgatg cctcctcttt tgaaagtgtg cgttgcacct tttgtgtgga 1380
tgccggggta tggtactatg aagtaacagt ggtcacttct ggcgtcatgc agattggctg 1440
ggccactcga gacagcaaat tcctcaatca tgaaggctac ggcattgggg atgatgaata 1500
ctcctgtgcg tatgatggct gccggcagct gatttggtac aatgccagaa gtaagcctca 1560
catacaccca tgctggaaag aaggagatac agtaggattt ctgttagact tgaatgaaaa 1620
gcaaatgatc ttctttttaa atggcaacca gctgcctcct gaaaagcaag tcttttcatc 1680
tactgtatct ggatttttg ctgcagctag tttcatgtca tatcaacaat gtgagttcaa 1740
ttttggagca aaaccattca aatacccacc atctatgaaa tttagcactt ttaatgacta 1800
cgccttccta acagctgaag aaaaaatcat tttgccaagg cacaggcgtc ttgctctgtt 1860
gaagcaagtc agtatccgag aaaactgctg ttccctttgt tgtgatgagg tagcagacac 1920
acaattgaag ccatgtggac acagtgacct gtgcatggat tgtgccttgc agctggagac 1980
ctgcccattg tgtcgtaaag aaatagtatc tagaatcaga cagatttctc atatttcatg 2040
acacatgtga agaggcatcg tggacttttt tctactcaat tccagccaat gttgaaaaga 2100
aaaagaaaaa aaaaactctc taatcagttg tacacacatt gaaacttata gccatggcca 2160
gattttatgc taaaaatggt agtttgtcaa agacaaaatt ctcttagaat ctaatccaac 2220
ttgccagccc tgagaaaatc ccttttaagg ccaaggaaag ctgaatgcta gcagccaggc 2280
ctgtggtact tccatgagaa accatagcag acaatgccct cccaagtact gaaatcacac 2340
tggaatcccc cttgttgggt tcatttgatt gtttaacaca ggatgtgttg tgtcattctg 2400
aagtttttat ttggggcaga agtctttatg gagatgtaaa tgacagcgtt tctgggttat 2460
gcataacttc tcactggtca gagacaccgg tgtgtcaagc atggatattg cattgcaaga 2520
cttgaatcta taaaaattag aatcacacag tcagtactac aagcaaaaca gagaacctga 2580
aagaaggtgc acagactgta agaaaaaacc caagtttgtg atatttcagt gattccaaag 2640
aacattctag gttttttgtt tgttttttg ttttttgggt ttttttttt tactgcagaa 2700

aattggtggt attttcacat tcatagtgtt tctatccaat ttcagtaccc acatttaatg 2760

aggaaaaaat gttttaccaa tgaaggagga attcttaaat tagctgtaat gttaggttgg 2820

agaaaatttg gtatttaggg tattttcaag gtaccatcaa atcagatttc tgtttttttg 2880

ttaaaaaaaa ttttttaat cagtattgtt tttacaagta atatactttg aaactcttga 2940

actaatagtc tcaaaaactc tagaggacag tctgagaaca cgtatttcta ttgttctaaa 3000

taaatacatg tttttgaata gttcaatcat gaattattga ctatgtcttc atcaaaagtg 3060

ttaatccctc tcagggtctc tggtgaagac cttcaagagt ttggtttttt ctcccaggaa 3120

attggaaggt agaattgtaa attcatagaa cttcttttat aatggtgtac ctcagcagct 3180

gcctttcaat ttatgccaag tccttacaga gtttatactt gaatagtaaa tatgtcttct 3240

gagttttaca gtgtcttaaa ctcaatgcac attttttttt cttctttttc caccccttct 3300

tgtttgtagt tcattatacc tgtcctatta cagaactgat ttccttcctg gctgtacatg 3360

ttggggtgct ggatttttt ccgtgtcttt agtcttccat aaatccacac acacacacac 3420

acacaaaaaa tatatatata tataaatata tatgtaggat acatgttctc ttctttagct 3480

tgtggtgaat acagtaattt gcattgaaga ataaaacatc tgttgccttt tttgactaag 3540

<210> 6
<211> 576
<212> PRT
<213> Homo sapiens

<400> 6

Met Ile Val Phe Gly Trp Ala Val Phe Leu Ala Ser Arg Ser Leu Gly
1 5 10 15

Gln Gly Leu Leu Leu Thr Leu Glu Glu His Ile Ala His Phe Leu Gly
20 25 30

Thr Gly Gly Ala Ala Thr Thr Met Gly Asn Ser Cys Ile Cys Arg Asp
35 40 45

Asp Ser Gly Thr Asp Asp Ser Val Asp Thr Gln Gln Gln Gln Ala Glu
50 55 60

Asn Ser Ala Val Pro Thr Ala Asp Thr Arg Ser Gln Pro Arg Asp Pro
65 70 75 80

Val Arg Pro Pro Arg Arg Gly Arg Gly Pro His Glu Pro Arg Arg Lys
85 90 95

Lys Gln Asn Val Asp Gly Leu Val Leu Asp Thr Leu Ala Val Ile Arg
100 105 110

```
Thr Leu Val Asp Asn Asp Gln Glu Pro Pro Tyr Ser Met Ile Thr Leu
        115                 120                 125

His Glu Met Ala Glu Thr Asp Glu Gly Trp Leu Asp Val Val Gln Ser
    130                 135                 140

Leu Ile Arg Val Ile Pro Leu Glu Asp Pro Leu Gly Pro Ala Val Ile
145                 150                 155                 160

Thr Leu Leu Leu Asp Glu Cys Pro Leu Pro Thr Lys Asp Ala Leu Gln
                165                 170                 175

Lys Leu Thr Glu Ile Leu Asn Leu Asn Gly Glu Val Ala Cys Gln Asp
            180                 185                 190

Ser Ser His Pro Ala Lys His Arg Asn Thr Ser Ala Val Leu Gly Cys
        195                 200                 205

Leu Ala Glu Lys Leu Ala Gly Pro Ala Ser Ile Gly Leu Leu Ser Pro
    210                 215                 220

Gly Ile Leu Glu Tyr Leu Leu Gln Cys Leu Lys Leu Gln Ser His Pro
225                 230                 235                 240

Thr Val Met Leu Phe Ala Leu Ile Ala Leu Glu Lys Phe Ala Gln Thr
                245                 250                 255

Ser Glu Asn Lys Leu Thr Ile Ser Glu Ser Ser Ile Ser Asp Arg Leu
            260                 265                 270

Val Thr Leu Glu Ser Trp Ala Asn Asp Pro Asp Tyr Leu Lys Arg Gln
        275                 280                 285

Val Gly Phe Cys Ala Gln Trp Ser Leu Asp Asn Leu Phe Leu Lys Glu
    290                 295                 300

Gly Arg Gln Leu Thr Tyr Glu Lys Val Asn Leu Ser Ser Ile Arg Ala
305                 310                 315                 320

Met Leu Asn Ser Asn Asp Val Ser Glu Tyr Leu Lys Ile Ser Pro His
                325                 330                 335

Gly Leu Glu Ala Arg Cys Asp Ala Ser Ser Phe Glu Ser Val Arg Cys
            340                 345                 350

Thr Phe Cys Val Asp Ala Gly Val Trp Tyr Tyr Glu Val Thr Val Val
        355                 360                 365
```

Thr Ser Gly Val Met Gln Ile Gly Trp Ala Thr Arg Asp Ser Lys Phe
370 375 380

Leu Asn His Glu Gly Tyr Gly Ile Gly Asp Asp Glu Tyr Ser Cys Ala
385 390 395 400

Tyr Asp Gly Cys Arg Gln Leu Ile Trp Tyr Asn Ala Arg Ser Lys Pro
405 410 415

His Ile His Pro Cys Trp Lys Glu Gly Asp Thr Val Gly Phe Leu Leu
420 425 430

Asp Leu Asn Glu Lys Gln Met Ile Phe Phe Leu Asn Gly Asn Gln Leu
435 440 445

Pro Pro Glu Lys Gln Val Phe Ser Ser Thr Val Ser Gly Phe Phe Ala
450 455 460

Ala Ala Ser Phe Met Ser Tyr Gln Gln Cys Glu Phe Asn Phe Gly Ala
465 470 475 480

Lys Pro Phe Lys Tyr Pro Pro Ser Met Lys Phe Ser Thr Phe Asn Asp
485 490 495

Tyr Ala Phe Leu Thr Ala Glu Glu Lys Ile Ile Leu Pro Arg His Arg
500 505 510

Arg Leu Ala Leu Leu Lys Gln Val Ser Ile Arg Glu Asn Cys Cys Ser
515 520 525

Leu Cys Cys Asp Glu Val Ala Asp Thr Gln Leu Lys Pro Cys Gly His
530 535 540

Ser Asp Leu Cys Met Asp Cys Ala Leu Gln Leu Glu Thr Cys Pro Leu
545 550 555 560

Cys Arg Lys Glu Ile Val Ser Arg Ile Arg Gln Ile Ser His Ile Ser
565 570 575

<210> 7
<211> 2589
<212> DNA
<213> Homo sapiens

<400> 7
ttggagacgg agtctcactc tgtcacccat tctggagtac aatggcatga tctcagctca 60

ctgcaacctc ctccttctgg gttcaagtga ttcttctgcc tcagcctcct gagtagctgg 120

gactacaggc ttgtgccacc acactcagct aatttttgta ttttcagtag agacggggtc 180

tccccatatt ggccaggctg gtcttgaact cctgacctca aacgatctgc ccaccccagc 240
ctcccaaagt gctgggatta caagtgtgag ccaccacacc cagcctaatg taactgtttt 300
cctttccttt tagtctttat ctcattgtta tatgtgatta taatgttgtc atttatgcat 360
tcacatcacc agtaattatc tcaaacagat cttaaatgct gcctatcatg ttgaagtcac 420
gtttcactca gggtcgtcag ttacaaggat gctatgggag caaatcaaac agatcatcca 480
gcgcatcaca tgggtgagcc cacctgccat cactctggaa tggaagagga aggtggccca 540
ggaagccatt gagagcctca gcgcctccaa attggctaag agcatttgca gccaattccg 600
gactcggctc aatagttccc acgaggcttt tgcagcctcc ttgcggcagc tggaagctgg 660
ccactcaggc cggttagaga aaacggaaga tctatggctg agggttcgga aagatcatgc 720
tccccgcctg gcccgccttt ctctggaaag ctgttcttta caggatgtct tgcttcatcg 780
taaacctaaa ctgggacagg aactgggccg gggccagtat ggtgtggtat acctgtgtga 840
caactgggga ggacacttcc cttgtgccct caaatcagtt gtccctccag atgagaagca 900
ctggaatgat ctggctttgg aatttcacta tatgaggtct ctgccgaagc atgagcgatt 960
ggtggatctc catggttcag tcattgacta caactatggt ggtggctcca gcattgctgt 1020
gctcctcatt atggagcggc tacaccggga tctctacaca gggctgaagg ctgggctgac 1080
cctggagaca cgtttgcaga tagcactaga tgtggtggag ggaatccgct tcctgcacag 1140
ccagggactt gtccatcgtg atatcaaact gaaaaatgtg ctgctggata agcagaaccg 1200
tgccaagatc actgacttag gattctgcaa gccagaggcc atgatgtcag gcagcattgt 1260
ggggacacca atccatatgg cccctgaact tttcacaggg aagtacgata attccgtgga 1320
tgtctacgct tttggaattc ttttctggta tatctgctca ggctctgtca agctccctga 1380
ggcatttgag aggtgtgcta gcaaagacca tctctggaac aatgtgcgga ggggggctcg 1440
cccagaacgt cttcctgtgt ttgatgagga gtgctggcag ttgatggaag cctgttggga 1500
tggcgacccc ttgaagaggc ctctcttggg cattgtccag cccatgctcc agggcatcat 1560
gaatcggctc tgcaagtcca attctgagca gccaaacaga ggactagatg attctacttg 1620
aaagcaaaga cctttctctt tcactctcta gttatttcct tccccctcac cttttggcca 1680
tggggagaat ttgacattta ttcactatag gacacactcc caagggaact ggtgcttgct 1740
gggaaacttg gaaccttccc aggcagggat gactcctgga cagtgaagag ttgaatgact 1800
gagcatattc agcagctcac tgaagcgcca agctatccct ttagcaaaaa agtgtctcag 1860
atgtgtaaaa gctgaggaat gtggtgttct ggcttcacaa atgaaaagga ggcagatgtt 1920
accattgtct tttcactgta tatacttcta agacagcaag cgggacactg cagtggcaat 1980
agtgttaaaa aatctcattc tcatgatttt tggctctagc taggaatagt ttagtcagga 2040
ctcagaaaat ttgagcttag ttctgggatt atgaaaacat ggggacaaaa acaataactt 2100

gtggatgtct ggttcctgct gtctgcagcc aggtatctca ggttgcaatg gtcagaagtc 2160

ccagtgaggg agctagaaac agagctatct gttcctcata gtgccagtgt gtttacattt 2220

ataggaccac atatccctgg tttctggggg agggtttcac tgttacatca ccaagaggcc 2280

ctgaaaagag gaaaaaaaaa aaaaaagagg ccaggcatgg tggctcacgc ctgtaatctc 2340

agcactttgg gaggctgagg cgggcggatc acaaggtcca gagttcaaga gcagcctggc 2400

caacatggtg aaaccccgtc tctactaaag atacaaaaaa ttagccgggc gtggtggtgc 2460

atgcctgtaa tccaagatac tcaggaggct gaggcaggag aatcgcttga acctgggagg 2520

cggaggttgc agtgagccga gatcatgcca ttgcactcca gcctgggcga cagagtgaga 2580

ctctgtctc 2589

<210> 8
<211> 390
<212> PRT
<213> Homo sapiens

<400> 8

Met Leu Trp Glu Gln Ile Lys Gln Ile Ile Gln Arg Ile Thr Trp Val
1 5 10 15

Ser Pro Pro Ala Ile Thr Leu Glu Trp Lys Arg Lys Val Ala Gln Glu
20 25 30

Ala Ile Glu Ser Leu Ser Ala Ser Lys Leu Ala Lys Ser Ile Cys Ser
35 40 45

Gln Phe Arg Thr Arg Leu Asn Ser Ser His Glu Ala Phe Ala Ala Ser
50 55 60

Leu Arg Gln Leu Glu Ala Gly His Ser Gly Arg Leu Glu Lys Thr Glu
65 70 75 80

Asp Leu Trp Leu Arg Val Arg Lys Asp His Ala Pro Arg Leu Ala Arg
85 90 95

Leu Ser Leu Glu Ser Cys Ser Leu Gln Asp Val Leu Leu His Arg Lys
100 105 110

Pro Lys Leu Gly Gln Glu Leu Gly Arg Gly Gln Tyr Gly Val Val Tyr
115 120 125

Leu Cys Asp Asn Trp Gly Gly His Phe Pro Cys Ala Leu Lys Ser Val
130 135 140

Val Pro Pro Asp Glu Lys His Trp Asn Asp Leu Ala Leu Glu Phe His

```
145                 150                 155                 160

Tyr Met Arg Ser Leu Pro Lys His Glu Arg Leu Val Asp Leu His Gly
                165                 170                 175

Ser Val Ile Asp Tyr Asn Tyr Gly Gly Gly Ser Ser Ile Ala Val Leu
            180                 185                 190

Leu Ile Met Glu Arg Leu His Arg Asp Leu Tyr Thr Gly Leu Lys Ala
        195                 200                 205

Gly Leu Thr Leu Glu Thr Arg Leu Gln Ile Ala Leu Asp Val Val Glu
    210                 215                 220

Gly Ile Arg Phe Leu His Ser Gln Gly Leu Val His Arg Asp Ile Lys
225                 230                 235                 240

Leu Lys Asn Val Leu Leu Asp Lys Gln Asn Arg Ala Lys Ile Thr Asp
                245                 250                 255

Leu Gly Phe Cys Lys Pro Glu Ala Met Met Ser Gly Ser Ile Val Gly
            260                 265                 270

Thr Pro Ile His Met Ala Pro Glu Leu Phe Thr Gly Lys Tyr Asp Asn
        275                 280                 285

Ser Val Asp Val Tyr Ala Phe Gly Ile Leu Phe Trp Tyr Ile Cys Ser
    290                 295                 300

Gly Ser Val Lys Leu Pro Glu Ala Phe Glu Arg Cys Ala Ser Lys Asp
305                 310                 315                 320

His Leu Trp Asn Asn Val Arg Arg Gly Ala Arg Pro Glu Arg Leu Pro
                325                 330                 335

Val Phe Asp Glu Glu Cys Trp Gln Leu Met Glu Ala Cys Trp Asp Gly
            340                 345                 350

Asp Pro Leu Lys Arg Pro Leu Leu Gly Ile Val Gln Pro Met Leu Gln
        355                 360                 365

Gly Ile Met Asn Arg Leu Cys Lys Ser Asn Ser Glu Gln Pro Asn Arg
    370                 375                 380

Gly Leu Asp Asp Ser Thr
385                 390
```

<210> 9
<211> 5766
<212> DNA
<213> Homo sapiens

<400> 9
cgccgccgca gccgccgcca ccaacgccaa caacaacgtg tcgggagcag ccgcggccga 60

agccggagca ggggccggag tcgcggctgg agtggggcct gcgcgtgtga gtggccgaga 120

gttgcccgtg ccggagacct caaactccgg gcgggagcgc tttcggtggg ccgagccgcc 180

gtcgcccgcg gcaccgtgtc cccaaggccc agacagcggg gcgggggtct atcaggcgcg 240

ggactggctt ctctcaggcc gacccttcta gccgggctcc ggatcttcct cggcctcggg 300

ccccttcac gttcaggacg gaggccggag aacggtgatc atcacacacc ccagccggct 360

tcctggccag ggcattttct ctgcggccaa gcagtgttgg gattgtagct gctgggagcc 420

ccctcctgcc gccacctctc ctgcaaggaa gggactgacg cgttcccact ttcttactct 480

gcacctcccc actctcttcc cattcaagaa cgagtttctc ttccctctcc attggaaatc 540

ctgctccgga ggtccaaaac acgtatttta agcgtgccct tcctcacctc ctggagcacc 600

cttagaagta ttgcctactt atccgggctg agaatccttc atttttgacc tggcttttt 660

ttcgcccttt gggagataaa ggtccctctc caccctctac taacactctg cacccaaggc 720

cttatccttt ggggtcacca gctccttggc catttctatg tgatttcccc cacccatctg 780

agttccagtt tcctctgggc tccaatctcc agtccctggc ggatctggtc agtcccaccc 840

ctaggtggga gtgaccaggg ggctctggcc caggattccc aaccctggaa ggcagctcca 900

aggcaggaag agaattgggc atcgggtacg aacctggcag ctcaggagtc ggggctccac 960

tcaccccaca caaaaagatg aaaaagcgca aagagctcaa tgcattgatt ggtttggctg 1020

gggacagccg gagaaagaag cccaagaaag gcccaagcag tcaccgcctg cttcgcactg 1080

agcctcccga ctcagactct gagtccagct ccgaagagga agaggaattc ggtgtggttg 1140

gaaatcgctc tcgctttgcc aagggagact atttacgatg ctgcaagatc tgttatccgc 1200

tctgtggttt tgtcatcctt gctgcctgtg ttgtggcctg tgttggcttg gtgtggatgc 1260

aggttgctct caaggaggat ctggatgccc tcaaggaaaa atttcgaaca atggaatcta 1320

atcagaaaag ctcattccaa gaaatcccca aacttaatga agaactactc agcaagcaaa 1380

aacaacttga gaagattgaa tctggagaga tgggtttgaa caaagtctgg ataaacatca 1440

cagaaatgaa taagcagatt tctctgttga cttctgcagt gaaccacctc aaagccaatg 1500

ttaagtcagc tgcagacttg attagcctgc ctaccactgt agagggactt cagaagagtg 1560

tagcttccat tggcaatact ttaaacagcg tccatcttgc tgtggaagca ctacagaaaa 1620

ctgtggatga acacaagaaa acgatggaat tactgcagag tgatatgaat cagcacttct 1680

tgaaggagac tcctggaagc aaccagatca ttccgtcacc ttcagccaca tcagaacttg 1740

acaataaaac ccacagtgag aatttgaaac aggatatcct gtaccttcac aactctttag 1800
aggaggtaaa cagtgcccta gtggggtacc agagacagaa tgatcttaaa ctcgagggaa 1860
tgaacgagac agtcagtaat cttacccaga gagtcaacct gatagaaagc gatgtggttg 1920
ctatgagcaa ggtagaaaag aaagcaaacc tgtccttcag catgatgggt gatagatctg 1980
ccactctgaa aagacagtct ttggatcaag tcaccaacag aacagataca gtaaaaatcc 2040
aaagcataaa gaaagaagat agttcaaatt ctcaggtatc caagctaaga gagaaactcc 2100
agctgatcag tgctcttaca aacaaacctg agagcaacag gcctccagag accgccgatg 2160
aagagcaagt agagagtttc acatcaaagc catcagcatt gccaaaattt tcacagtttc 2220
ttggagaccc agttgagaaa gctgcccaac taagacctat ctccctacca ggagtttcta 2280
gcactgaaga tcttcaggat ttattccgca agactggcca ggacgtggat gggaagctga 2340
cctaccagga aatctggacc tccctaggtt ctgctatgcc agaaccagag agcttgagag 2400
catttgattc cgatggagat ggaagatact cattcctgga gctaagggta gctttaggta 2460
tctagcttca tcaggcatat tttagaaatg gactgcctaa tatctattta cctaacaaca 2520
aaacaaccct tacttaccca tcagtcctct agtcctccaa actactgtag cagatacttt 2580
gccacctttt aacttgtttg aagaagctat ataaaagtta ttttttaaa gaagaagacc 2640
attttactta tgatgttcag aaatctatga tttcctacaa ccagtaagat cttacatttt 2700
aaaattgcca gaaaaaaaat taaagccctc tttttttctc tttccttttt ttgaggggag 2760
gagaccttat cttttaaagc tgggaaatgt atatagagag agaataagcc acttttatat 2820
ttcacttaaa tttgccttaa attagctgca ctttatagag actcagaaaa tgtcttttct 2880
ttaaaagata ggccttttct gtttgtaaat atttaaatga aagaaagcat tgtgcatatt 2940
gtgtggaaag taggaagaat ggttttgaac aggatatgaa caaatgactt attaaaaatt 3000
gctgatctgg tgtaggtggc agctgaaact acatccatgt ctccataagg catccctcaa 3060
aggcccaggc gctgccaggg ggtttgtcct ggtagctgga ggaaccgatt tcagggaata 3120
gacactggag acaatactga ctccaggcat ggctcatgga agtaggattc tggttctttg 3180
ttcctattcc ctcagctaat cccaacctgg gaatcagaga agtcttgggg atttttctca 3240
tttttagtac tatttcaggg tttatgagca taaaaagtta tccattgggg agctccattt 3300
tccctgctga gtgagctaga ttgccttccc cacccaccca cttaagtctg tcttaaagcc 3360
gtagctggct cccaccacca gtaccatctc catttgaatg gcagggctaa attccccccag 3420
ccattatctc acactgacca cccagagctt tagaagagag ctgtgcttct aattttgacc 3480
cagaaaacca tacccccttga gattttacct agaggctaac caagagccta atatgtttct 3540
ctgggggatg actaaagcca aaaaggctgt gagatgaaac atgtgaaata atattcagtt 3600
tccttaccat taccagctca gaagtagcta gaggctttct acccaaagga tgccaaagta 3660

tagcagggca ggcctggagc tagggccttc acatggtggt agcaagtttt tcaaatctaa 3720
tacaatcaag tacaatactt cctttaaatg cttctgtgga cctggcatga aagatcccta 3780
gattgaaagg aataatacct ccatgtctcc tgtatgttga gtctagaatt gctgtgttgt 3840
tcttagaagc agtctttggg caacaacttg aaaggggaaa aaaaaactac aaaaacttaa 3900
ctttggtata ggccaagtca gggagaaagt agagaaagct gtcatgccac agacttcttt 3960
agtggagatc atttcctttt taactttgtt caggttgccc ttcaccatgg atacagtccg 4020
gtacccttaa acatttaagg gctgtttttt tttttcttta tatgatgttc agcttggtat 4080
taaccaaact taaatttttt ttccagaagt attaaaattt agttaaagca agatgaagtc 4140
ttttcccatg aatgtgtccc cttatctcat tatagcttat gctcagtttc actttcttgg 4200
aaaggttaaa acaaattagc ctggcacttt aggtaacttg aaaataactc tcacccttct 4260
gactgccttc catccactct caccccacac ctttttttaa tatatataca cccttacaga 4320
ttttctaaca accaaataaa attctagcaa taaattgaat tatactgcta tatttgtata 4380
tactgtacca taaatagtat gtctgtacct ggaaggtatt tttttgagaa ctgatttata 4440
tgaaatatac ttgagggtaa tgtagcttgt ccttttttagt ttcaaattct tattcatagg 4500
cagatacttt gaagacacct taactctttg ttgtgtgtat tttcccttgc atcagataac 4560
tatacatgga gaaatccatt ttgttttttg gtttatcttc tgttcttcct accttgtgct 4620
tgtctgacta gtatcacccc tgagccagta tacagtacca tcctctcccc ctaccaagtt 4680
tggcagaggt gtttatactg cattctcaga tctatgaata tttttgtcat cctgcctgag 4740
aaagtacatc ttccaggcaa aagtagggta tcttgaactc ctttcttctg agttgagcag 4800
actcactggc ttaagcatcc cataagccta ttatttagtg tgaatttggg acttttattt 4860
gtatgttgtc ttaacttact gagcctgtga cttgtattta gctgtagtca agcctcacca 4920
gccatatctc atctctgaga ggaagactgg atagtgagtt tggttttggt tttggtttta 4980
aacccttta ccttttcagt ctggcagctc ctactgctgc tttcctaagt tactcaaagc 5040
actgcttctt ggtccttgtt ttcttttgcc ttgtactatc ttgtataatg tttactaaaa 5100
gagcctcccc ataaatcaga gaggcaaaaa ggtagcatat aaaaagaaga accaaaagaa 5160
gagaaaacct gacttctaga tgactgtggt gtcatctaaa aggcagaggt gcctgaccct 5220
caggaatgtg tatgatgcaa cagagctgaa gaatgggttc ccttcaccta ctgctggcca 5280
ttgtcccttc tgttggagat tcctttcttt atggttgctg ttcataaccc cttcaacacc 5340
ccccttgct ggtcctgccg gtcacagttg aggattttgg ctgtgatggg ctcatactca 5400
taatggcctg ttgagttgtt tttagtgaca ggttttgttt gctgcacacc tcattggctt 5460
gcttaccaac accagctaga agtggtccct ctttcatatc caaaccagaa cacatttggg 5520
tattcctgag actttataga gcgtgtattg tttgttatac gaacctaacc taccttgttg 5580

ttttctcata ttaagaaatt taaatctact ttgttttagt gaagctattt tggtaatgta 5640

taagcaattt tggttctgtt tagggagaat tcagggtttc catgacttga tttgagcttt 5700

gcatatttca ataagtaac acagccattt taatttagct aaataaaaaa atttcttttt 5760

tctctc 5766

<210> 10
<211> 495
<212> PRT
<213> Homo sapiens

<400> 10

Met Lys Lys Arg Lys Glu Leu Asn Ala Leu Ile Gly Leu Ala Gly Asp
1 5 10 15

Ser Arg Arg Lys Lys Pro Lys Lys Gly Pro Ser Ser His Arg Leu Leu
20 25 30

Arg Thr Glu Pro Pro Asp Ser Asp Ser Glu Ser Ser Ser Glu Glu Glu
35 40 45

Glu Glu Phe Gly Val Val Gly Asn Arg Ser Arg Phe Ala Lys Gly Asp
50 55 60

Tyr Leu Arg Cys Cys Lys Ile Cys Tyr Pro Leu Cys Gly Phe Val Ile
65 70 75 80

Leu Ala Ala Cys Val Val Ala Cys Val Gly Leu Val Trp Met Gln Val
85 90 95

Ala Leu Lys Glu Asp Leu Asp Ala Leu Lys Glu Lys Phe Arg Thr Met
100 105 110

Glu Ser Asn Gln Lys Ser Ser Phe Gln Glu Ile Pro Lys Leu Asn Glu
115 120 125

Glu Leu Leu Ser Lys Gln Lys Gln Leu Glu Lys Ile Glu Ser Gly Glu
130 135 140

Met Gly Leu Asn Lys Val Trp Ile Asn Ile Thr Glu Met Asn Lys Gln
145 150 155 160

Ile Ser Leu Leu Thr Ser Ala Val Asn His Leu Lys Ala Asn Val Lys
165 170 175

Ser Ala Ala Asp Leu Ile Ser Leu Pro Thr Thr Val Glu Gly Leu Gln
180 185 190

```
Lys Ser Val Ala Ser Ile Gly Asn Thr Leu Asn Ser Val His Leu Ala
        195                 200                 205

Val Glu Ala Leu Gln Lys Thr Val Asp Glu His Lys Lys Thr Met Glu
    210                 215                 220

Leu Leu Gln Ser Asp Met Asn Gln His Phe Leu Lys Glu Thr Pro Gly
225                 230                 235                 240

Ser Asn Gln Ile Ile Pro Ser Pro Ser Ala Thr Ser Glu Leu Asp Asn
                245                 250                 255

Lys Thr His Ser Glu Asn Leu Lys Gln Asp Ile Leu Tyr Leu His Asn
            260                 265                 270

Ser Leu Glu Glu Val Asn Ser Ala Leu Val Gly Tyr Gln Arg Gln Asn
        275                 280                 285

Asp Leu Lys Leu Glu Gly Met Asn Glu Thr Val Ser Asn Leu Thr Gln
    290                 295                 300

Arg Val Asn Leu Ile Glu Ser Asp Val Val Ala Met Ser Lys Val Glu
305                 310                 315                 320

Lys Lys Ala Asn Leu Ser Phe Ser Met Met Gly Asp Arg Ser Ala Thr
                325                 330                 335

Leu Lys Arg Gln Ser Leu Asp Gln Val Thr Asn Arg Thr Asp Thr Val
            340                 345                 350

Lys Ile Gln Ser Ile Lys Lys Glu Asp Ser Ser Asn Ser Gln Val Ser
        355                 360                 365

Lys Leu Arg Glu Lys Leu Gln Leu Ile Ser Ala Leu Thr Asn Lys Pro
    370                 375                 380

Glu Ser Asn Arg Pro Pro Glu Thr Ala Asp Glu Glu Gln Val Glu Ser
385                 390                 395                 400

Phe Thr Ser Lys Pro Ser Ala Leu Pro Lys Phe Ser Gln Phe Leu Gly
                405                 410                 415

Asp Pro Val Glu Lys Ala Ala Gln Leu Arg Pro Ile Ser Leu Pro Gly
            420                 425                 430

Val Ser Ser Thr Glu Asp Leu Gln Asp Leu Phe Arg Lys Thr Gly Gln
        435                 440                 445
```

Asp Val Asp Gly Lys Leu Thr Tyr Gln Glu Ile Trp Thr Ser Leu Gly
450 455 460

Ser Ala Met Pro Glu Pro Glu Ser Leu Arg Ala Phe Asp Ser Asp Gly
465 470 475 480

Asp Gly Arg Tyr Ser Phe Leu Glu Leu Arg Val Ala Leu Gly Ile
485 490 495

<210> 11
<211> 6550
<212> DNA
<213> Homo sapiens

<400> 11
tcccggaagg caaatggctg gcgtggaagc acaacccgct ttcactcttc gaatttgtgc 60

ttagctcttt tcttgtacct tgcgactcgt gaccaacatg ctgtgatgtg tgccgaggga 120

ggaattggtc agctacacaa cctggatctt accacagttt ggatatgact gaggctctcc 180

aatgggccag atatcactgg cgacggctga tcagaggtgc aaccagggat gatgattcag 240

ggccatacaa ctattcctcg ttgctcgcct gtgggcgcaa gtcctctcag acccctaaac 300

tgtcaggaag gcaccggatt gttgttcccc acatccagcc cttcaaggat gagtatgaga 360

agttctccgg agcctatgtg aacaatcgaa tacgaacaac aaagtacaca cttctgaatt 420

ttgtgccaag aaatttattt gaacaatttc acagagctgc caatttatat ttcctgttcc 480

tagttgtcct gaactgggta cctttggtag aagccttcca aaaggaaatc accatgttgc 540

ctctggtggt ggtccttaca attatcgcaa ttaaagatgg cctggaaggt tatcggaaat 600

acaaaattga caaacagatc aataatttaa taactaaagt ttatagtagg aaagagaaaa 660

aatacattga ccgatgctgg aaagacgtta ctgttgggga ctttattcgc ctctcctgca 720

acgaggtcat ccctgcagac atggtactac tcttttccac tgatccagat ggaatctgtc 780

acattgagac ttctggtctt gatggagaga gcaatttaaa acagaggcag gtggttcggg 840

gatatgcaga acaggactct gaagttgatc ctgagaagtt ttccagtagg atagaatgtg 900

aaagcccaaa caatgacctc agcagattcc gaggcttcct agaacattcc aacaaagaac 960

gcgtgggtct cagtaaagaa aatttgttgc ttagaggatg caccattaga aacacagagg 1020

ctgttgtggg cattgtggtt tatgcaggcc atgaaaccaa agcaatgctg aacaacagtg 1080

ggccacggta taagcgcagc aaattagaaa gaagagcaaa cacagatgtc ctctggtgtg 1140

tcatgcttct ggtcataatg tgcttaactg gcgcagtagg tcatggaatc tggctgagca 1200

ggtatgaaaa gatgcatttt ttcaatgttc ccgagcctga tggacatatc atatcaccac 1260

tgttggcagg attttatatg ttttggacca tgatcatttt gttacaggtc ttgattccta 1320

tttctctcta tgtttccatc gaaattgtga agcttggaca aatatatttc attcaaagtg 1380

atgtggattt ctacaatgaa aaaatggatt ctattgttca gtgccgagcc ctgaacatcg 1440
ccgaggatct gggacagatt cagtacctct tttccgataa gacaggaacc ctcactgaga 1500
ataagatggt ttttcgaaga tgtagtgtgg caggatttga ttactgccat gaagaaaatg 1560
ccaggaggtt ggagtcctat caggaagctg tctctgaaga tgaagatttt atagacacag 1620
tcagtggttc cctcagcaat atggcaaaac cgagagcccc cagctgcagg acagttcata 1680
atgggccttt gggaaataag ccctcaaatc atcttgctgg gagctctttt actctaggaa 1740
gtggagaagg agccagtgaa gtgcctcatt ccagacaggc tgctttcagt agccccattg 1800
aaacagacgt ggtaccagac accaggcttt tagacaaatt tagtcagatt acacctcggc 1860
tctttatgcc actagatgag accatccaaa atccaccaat ggaaactttg tacattatcg 1920
acttttttcat tgcattggca atttgcaaca cagtagtggt ttctgctcct aaccaacccc 1980
gacaaaagat cagacaccct tcactggggg ggttgcccat taagtctttg gaagagatta 2040
aaagtctttt ccagagatgg tctgtccgaa gatcaagttc tccatcgctt aacagtggga 2100
aagagccatc ttctggagtt ccaaacgcct ttgtgagcag actccctctc tttagtcgaa 2160
tgaaaccagc ttcacctgtg gaggaagagg tctcccaggt gtgtgagagc ccccagtgct 2220
ccagtagctc agcttgctgc acagaaacag agaaacaaca cggtgatgca ggcctcctga 2280
atggcaaggc agagtccctc cctggacagc cattggcctg caacctgtgt tatgaggccg 2340
agagcccaga cgaagcggcc ttagtgtatg ccgccagggc ttaccaatgc actttacggt 2400
ctcggacacc agagcaggtc atggtggact ttgctgcttt gggaccatta acatttcaac 2460
tcctacacat cctgcccttt gactcagtaa gaaaaagaat gtctgttgtg gtccgacacc 2520
ctctttccaa tcaagttgtg gtgtatacga aaggcgctga ttctgtgatc atggagttac 2580
tgtcggtggc ttccccagat ggagcaagtc tggagaaaca acagatgata gtaagggaga 2640
aaacccagaa gcacttggat gactatgcca aacaaggcct tcgtacttta tgtatagcaa 2700
agaaggtcat gagtgacact gaatatgcag agtggctgag gaatcatttt ttagctgaaa 2760
ccagcattga caacagggaa gaattactac ttgaatctgc catgaggttg gagaacaaac 2820
ttacattact tggtgctact ggcattgaag accgtctgca ggagggagtc cctgaatcta 2880
tagaagctct tcacaaagcg ggcatcaaga tctggatgct gacaggggac aagcaggaga 2940
cagctgtcaa catagcttat gcatgcaaac tactggagcc agatgacaag ctttttatcc 3000
tcaataccca aagtaaagat gcctgtggga tgctgatgag cacaattttg aaagaacttc 3060
agaagaaaac tcaagccctg ccagagcaag tgtcattaag tgaagattta cttcagcctc 3120
ctgtcccccg ggactcaggg ttacgagctg gactcattat cactgggagg accctggagt 3180
ttgccctgca agaaagtctg caaaagcagt tcctggaact gacatcttgg tgtcaagctg 3240
tggtctgctg ccgagccaca ccgctgcaga aaagtgaagt ggtgaaattg gtccgcagcc 3300

atctccaggt gatgaccctt gctattggtg atggtgccaa tgatgttagc atgatacaag 3360
tggcagacat tgggataggg gtctcaggtc aagaaggcat gcaggctgtg atggccagtg 3420
actttgccgt ttatcagttc aaacatctca gcaagctcct tcttgtccat ggacactggt 3480
gttatacacg gctttccaac atgattctct atttttttcta taagaatgtg gcctatgtga 3540
acctcctttt ctggtaccag ttcttttgtg gattttcagg gacatccatg actgattact 3600
gggttttgat cttcttcaac ctcctcttca catctgcccc tcctgtcatt tatggtgttt 3660
tggagaaaga tgtgtctgca gagaccctca tgcaactgcc tgaactttac agaagtggtc 3720
agaaatcaga ggcatactta ccccatacct tctggatcac cttattggat gctttttatc 3780
aaagcctggt ctgcttcttt gtgccttatt ttacctacca gggctcagat actgacatct 3840
ttgcatttgg aaaccccctg aacacagccg ctctgttcat cgttctcctc catctggtca 3900
ttgaaagcaa gagtttgact tggattcact tgctggtcat cattggtagc atcttgtctt 3960
atttttatt tgccatagtt tttggagcca tgtgtgtaac ttgcaaccca ccatccaacc 4020
cttactggat tatgcaggag cacatgctgg atccagtatt ctacttagtt tgtatcctca 4080
cgacgtccat tgctcttctg cccaggtttg tatacagagt tcttcaggga tccctgtttc 4140
catctccaat tctgagagct aagcactttg acagactaac tccagaggag aggactaaag 4200
ctctcaagaa gtggagaggg gctggaaaga tgaatcaagt gacatcaaag tatgctaacc 4260
aatcagctgg caagtcagga agaagaccca tgcctggccc ttctgctgta tttgcaatga 4320
agtcagcaag ttcctgtgct attgagcaag gaaacttatc tctgtgtgag actgctttag 4380
atcaaggcta ctctgaaact aaggcctttg agatggctgg accctccaaa ggtaaagaaa 4440
gctagatacc ctccttggag ttgcaagtat tctttcaagg ttggaagagg gattttgaag 4500
aggtatctct ccaagcaaga atgacttgtt tttccataag ggacatgagc attttactag 4560
gcttggaaga gctgacatga tgagcattat tgtatgtttg tatatacatt tgtgatagag 4620
ggctagagtt tgacctagag agagtttaag gaagtgaaat atttaattca gaaccaaatg 4680
cttttgtaaa acttttgga ttttgtaaaa gcattttcat tctcttagaa attcaagtat 4740
tttcaagggg agtcatttga gatatattta ttttactagg agatcttata ttctagggaa 4800
atgctttaaa tggtcaggct ccagtcggaa tttttttaag aaaaaagtag tttttaatac 4860
attggttagg actcagagga aatacggaaa aaacattgta gatggttaat ttacagataa 4920
aatcccaaga gccttttaaa caacaaggta cctaaaatag ggtataatta tactgcttaa 4980
aatacagtta gtgcctatta atagcttttt atttcctatg ggaagatgct tttggtcttc 5040
tggctgaaga tgtaggcata cctctcactc atttcaatgt tttcctgagg tggagccttc 5100
attggaaagg ggaaagaggg attctagggt ttcatcaggg accaggaatg cattcctctg 5160
tcaggttcca atcaagagaa gacctttttat gagatctgcc tctgtataga tgttgtcagt 5220

taggaaactg aagccatagg tcaggcagac atcagctcag cctgtggccc attgggtgat 5280

ttcctgtatt ttaaaactga cagtagcctg atcaaagtga tacaatcaat ttcaaaacaa 5340

tcttccagag accacttgaa ggttcatagt ttttacaata ccctgagact tttcaggtgt 5400

tggagcctct aaaatatgag atataaacag aaactaatac aagttgttct ctggaggttt 5460

ctatgaggtt cttagaaaaa tttggtttta aaatcatttg aggacaggaa tgtctatagc 5520

aagtttactc ctattgcgaa tcatgtatgc tggctttagt tgtaacaaac gattttattc 5580

taagtaaggc caggtgctac tataaaatca tatattcctg ttgaagttct tttgaagtcc 5640

tatctctatt tattatattt gaaagttgtc agccaccagt catccagaat ttccttcctg 5700

aatctccatg ctcatatgca atgtctacat caaggtcttc ttaatgacta ttattctcag 5760

ggtttagttt tctaccttct gcctactatt ttggtctgac atttttgtag ccttctgtta 5820

ttattggaaa tagtctctta cataagctga tttcgagaac tttcaaaatc tcacatagct 5880

aatggaagtt gctttctgct ttcttatgac tgtttttata aataaactgt ttcataaata 5940

atgtatactg gaaaattgtc atcactcttg taaacatttg ccaaatccaa ctgtaattag 6000

taccagaact tagaaggaaa gtcacaaagc atgcctttta ccagaatatt gctttgcctt 6060

ttcttatatg gaaaatcact gatgtcatat catttattat cctcacagca gctggcattc 6120

attccctgaa atcgtcgtct cctgataaac aataatttag agatatgttc acttgtaaat 6180

aaaaaggtta gaatctaagg acaaagaatg atccaaaaat actaaggaaa aagaacaatt 6240

ttcaattttc agaaatgaat tatctgtaca tatgaataga ggaaaatact gagtgctaat 6300

gttacagggc cacaatctca tatccacaat tccaaaatct ccaaacctct gaaaactgaa 6360

atatttttca taaattaggc gtaaattcta tgccctgaac aaatgtaaag ctatttatgt 6420

actttgctta atgtaatcat tcatatactt tgctacaaaa atattaatat atttcattac 6480

tgaatgctaa actgtatttc tttttaaaat ttggaaaatt tttaattctg aaacacaatc 6540

tagaccttag 6550

<210> 12
<211> 1426
<212> PRT
<213> Homo sapiens

<400> 12

Met Thr Glu Ala Leu Gln Trp Ala Arg Tyr His Trp Arg Arg Leu Ile
1 5 10 15

Arg Gly Ala Thr Arg Asp Asp Asp Ser Gly Pro Tyr Asn Tyr Ser Ser
20 25 30

Leu Leu Ala Cys Gly Arg Lys Ser Ser Gln Thr Pro Lys Leu Ser Gly

35 40 45

Arg His Arg Ile Val Val Pro His Ile Gln Pro Phe Lys Asp Glu Tyr
50 55 60

Glu Lys Phe Ser Gly Ala Tyr Val Asn Asn Arg Ile Arg Thr Thr Lys
65 70 75 80

Tyr Thr Leu Leu Asn Phe Val Pro Arg Asn Leu Phe Glu Gln Phe His
85 90 95

Arg Ala Ala Asn Leu Tyr Phe Leu Phe Leu Val Val Leu Asn Trp Val
100 105 110

Pro Leu Val Glu Ala Phe Gln Lys Glu Ile Thr Met Leu Pro Leu Val
115 120 125

Val Val Leu Thr Ile Ile Ala Ile Lys Asp Gly Leu Glu Gly Tyr Arg
130 135 140

Lys Tyr Lys Ile Asp Lys Gln Ile Asn Asn Leu Ile Thr Lys Val Tyr
145 150 155 160

Ser Arg Lys Glu Lys Lys Tyr Ile Asp Arg Cys Trp Lys Asp Val Thr
165 170 175

Val Gly Asp Phe Ile Arg Leu Ser Cys Asn Glu Val Ile Pro Ala Asp
180 185 190

Met Val Leu Leu Phe Ser Thr Asp Pro Asp Gly Ile Cys His Ile Glu
195 200 205

Thr Ser Gly Leu Asp Gly Glu Ser Asn Leu Lys Gln Arg Gln Val Val
210 215 220

Arg Gly Tyr Ala Glu Gln Asp Ser Glu Val Asp Pro Glu Lys Phe Ser
225 230 235 240

Ser Arg Ile Glu Cys Glu Ser Pro Asn Asn Asp Leu Ser Arg Phe Arg
245 250 255

Gly Phe Leu Glu His Ser Asn Lys Glu Arg Val Gly Leu Ser Lys Glu
260 265 270

Asn Leu Leu Leu Arg Gly Cys Thr Ile Arg Asn Thr Glu Ala Val Val
275 280 285

Gly Ile Val Val Tyr Ala Gly His Glu Thr Lys Ala Met Leu Asn Asn
290 295 300

Ser Gly Pro Arg Tyr Lys Arg Ser Lys Leu Glu Arg Arg Ala Asn Thr
305 310 315 320

Asp Val Leu Trp Cys Val Met Leu Leu Val Ile Met Cys Leu Thr Gly
325 330 335

Ala Val Gly His Gly Ile Trp Leu Ser Arg Tyr Glu Lys Met His Phe
340 345 350

Phe Asn Val Pro Glu Pro Asp Gly His Ile Ile Ser Pro Leu Leu Ala
355 360 365

Gly Phe Tyr Met Phe Trp Thr Met Ile Ile Leu Leu Gln Val Leu Ile
370 375 380

Pro Ile Ser Leu Tyr Val Ser Ile Glu Ile Val Lys Leu Gly Gln Ile
385 390 395 400

Tyr Phe Ile Gln Ser Asp Val Asp Phe Tyr Asn Glu Lys Met Asp Ser
405 410 415

Ile Val Gln Cys Arg Ala Leu Asn Ile Ala Glu Asp Leu Gly Gln Ile
420 425 430

Gln Tyr Leu Phe Ser Asp Lys Thr Gly Thr Leu Thr Glu Asn Lys Met
435 440 445

Val Phe Arg Arg Cys Ser Val Ala Gly Phe Asp Tyr Cys His Glu Glu
450 455 460

Asn Ala Arg Arg Leu Glu Ser Tyr Gln Glu Ala Val Ser Glu Asp Glu
465 470 475 480

Asp Phe Ile Asp Thr Val Ser Gly Ser Leu Ser Asn Met Ala Lys Pro
485 490 495

Arg Ala Pro Ser Cys Arg Thr Val His Asn Gly Pro Leu Gly Asn Lys
500 505 510

Pro Ser Asn His Leu Ala Gly Ser Ser Phe Thr Leu Gly Ser Gly Glu
515 520 525

Gly Ala Ser Glu Val Pro His Ser Arg Gln Ala Ala Phe Ser Ser Pro
530 535 540

```
Ile Glu Thr Asp Val Val Pro Asp Thr Arg Leu Leu Asp Lys Phe Ser
545                 550                 555                 560

Gln Ile Thr Pro Arg Leu Phe Met Pro Leu Asp Glu Thr Ile Gln Asn
                565                 570                 575

Pro Pro Met Glu Thr Leu Tyr Ile Ile Asp Phe Phe Ile Ala Leu Ala
            580                 585                 590

Ile Cys Asn Thr Val Val Val Ser Ala Pro Asn Gln Pro Arg Gln Lys
        595                 600                 605

Ile Arg His Pro Ser Leu Gly Gly Leu Pro Ile Lys Ser Leu Glu Glu
    610                 615                 620

Ile Lys Ser Leu Phe Gln Arg Trp Ser Val Arg Arg Ser Ser Ser Pro
625                 630                 635                 640

Ser Leu Asn Ser Gly Lys Glu Pro Ser Ser Gly Val Pro Asn Ala Phe
                645                 650                 655

Val Ser Arg Leu Pro Leu Phe Ser Arg Met Lys Pro Ala Ser Pro Val
            660                 665                 670

Glu Glu Glu Val Ser Gln Val Cys Glu Ser Pro Gln Cys Ser Ser Ser
        675                 680                 685

Ser Ala Cys Cys Thr Glu Thr Glu Lys Gln His Gly Asp Ala Gly Leu
    690                 695                 700

Leu Asn Gly Lys Ala Glu Ser Leu Pro Gly Gln Pro Leu Ala Cys Asn
705                 710                 715                 720

Leu Cys Tyr Glu Ala Glu Ser Pro Asp Glu Ala Ala Leu Val Tyr Ala
                725                 730                 735

Ala Arg Ala Tyr Gln Cys Thr Leu Arg Ser Arg Thr Pro Glu Gln Val
            740                 745                 750

Met Val Asp Phe Ala Ala Leu Gly Pro Leu Thr Phe Gln Leu Leu His
        755                 760                 765

Ile Leu Pro Phe Asp Ser Val Arg Lys Arg Met Ser Val Val Val Arg
    770                 775                 780

His Pro Leu Ser Asn Gln Val Val Val Tyr Thr Lys Gly Ala Asp Ser
785                 790                 795                 800
```

```
Val Ile Met Glu Leu Leu Ser Val Ala Ser Pro Asp Gly Ala Ser Leu
                805                 810                 815

Glu Lys Gln Gln Met Ile Val Arg Glu Lys Thr Gln Lys His Leu Asp
            820                 825                 830

Asp Tyr Ala Lys Gln Gly Leu Arg Thr Leu Cys Ile Ala Lys Lys Val
        835                 840                 845

Met Ser Asp Thr Glu Tyr Ala Glu Trp Leu Arg Asn His Phe Leu Ala
    850                 855                 860

Glu Thr Ser Ile Asp Asn Arg Glu Glu Leu Leu Leu Glu Ser Ala Met
865                 870                 875                 880

Arg Leu Glu Asn Lys Leu Thr Leu Leu Gly Ala Thr Gly Ile Glu Asp
                885                 890                 895

Arg Leu Gln Glu Gly Val Pro Glu Ser Ile Glu Ala Leu His Lys Ala
            900                 905                 910

Gly Ile Lys Ile Trp Met Leu Thr Gly Asp Lys Gln Glu Thr Ala Val
        915                 920                 925

Asn Ile Ala Tyr Ala Cys Lys Leu Leu Glu Pro Asp Asp Lys Leu Phe
    930                 935                 940

Ile Leu Asn Thr Gln Ser Lys Asp Ala Cys Gly Met Leu Met Ser Thr
945                 950                 955                 960

Ile Leu Lys Glu Leu Gln Lys Lys Thr Gln Ala Leu Pro Glu Gln Val
                965                 970                 975

Ser Leu Ser Glu Asp Leu Leu Gln Pro Pro Val Pro Arg Asp Ser Gly
            980                 985                 990

Leu Arg Ala Gly Leu Ile Ile Thr  Gly Arg Thr Leu Glu  Phe Ala Leu
        995                 1000                1005

Gln Glu  Ser Leu Gln Lys Gln  Phe Leu Glu Leu Thr  Ser Trp Cys
    1010                1015                1020

Gln Ala  Val Val Cys Cys Arg  Ala Thr Pro Leu Gln  Lys Ser Glu
    1025                1030                1035

Val Val  Lys Leu Val Arg Ser  His Leu Gln Val Met  Thr Leu Ala
    1040                1045                1050
```

Ile Gly Asp Gly Ala Asn Asp Val Ser Met Ile Gln Val Ala Asp
1055 1060 1065

Ile Gly Ile Gly Val Ser Gly Gln Glu Gly Met Gln Ala Val Met
1070 1075 1080

Ala Ser Asp Phe Ala Val Tyr Gln Phe Lys His Leu Ser Lys Leu
1085 1090 1095

Leu Leu Val His Gly His Trp Cys Tyr Thr Arg Leu Ser Asn Met
1100 1105 1110

Ile Leu Tyr Phe Phe Tyr Lys Asn Val Ala Tyr Val Asn Leu Leu
1115 1120 1125

Phe Trp Tyr Gln Phe Phe Cys Gly Phe Ser Gly Thr Ser Met Thr
1130 1135 1140

Asp Tyr Trp Val Leu Ile Phe Phe Asn Leu Leu Phe Thr Ser Ala
1145 1150 1155

Pro Pro Val Ile Tyr Gly Val Leu Glu Lys Asp Val Ser Ala Glu
1160 1165 1170

Thr Leu Met Gln Leu Pro Glu Leu Tyr Arg Ser Gly Gln Lys Ser
1175 1180 1185

Glu Ala Tyr Leu Pro His Thr Phe Trp Ile Thr Leu Leu Asp Ala
1190 1195 1200

Phe Tyr Gln Ser Leu Val Cys Phe Phe Val Pro Tyr Phe Thr Tyr
1205 1210 1215

Gln Gly Ser Asp Thr Asp Ile Phe Ala Phe Gly Asn Pro Leu Asn
1220 1225 1230

Thr Ala Ala Leu Phe Ile Val Leu Leu His Leu Val Ile Glu Ser
1235 1240 1245

Lys Ser Leu Thr Trp Ile His Leu Leu Val Ile Ile Gly Ser Ile
1250 1255 1260

Leu Ser Tyr Phe Leu Phe Ala Ile Val Phe Gly Ala Met Cys Val
1265 1270 1275

Thr Cys Asn Pro Pro Ser Asn Pro Tyr Trp Ile Met Gln Glu His
1280 1285 1290

Met Leu Asp Pro Val Phe Tyr Leu Val Cys Ile Leu Thr Thr Ser
1295 1300 1305

Ile Ala Leu Leu Pro Arg Phe Val Tyr Arg Val Leu Gln Gly Ser
1310 1315 1320

Leu Phe Pro Ser Pro Ile Leu Arg Ala Lys His Phe Asp Arg Leu
1325 1330 1335

Thr Pro Glu Glu Arg Thr Lys Ala Leu Lys Lys Trp Arg Gly Ala
1340 1345 1350

Gly Lys Met Asn Gln Val Thr Ser Lys Tyr Ala Asn Gln Ser Ala
1355 1360 1365

Gly Lys Ser Gly Arg Arg Pro Met Pro Gly Pro Ser Ala Val Phe
1370 1375 1380

Ala Met Lys Ser Ala Ser Ser Cys Ala Ile Glu Gln Gly Asn Leu
1385 1390 1395

Ser Leu Cys Glu Thr Ala Leu Asp Gln Gly Tyr Ser Glu Thr Lys
1400 1405 1410

Ala Phe Glu Met Ala Gly Pro Ser Lys Gly Lys Glu Ser
1415 1420 1425

<210> 13
<211> 8083
<212> DNA
<213> Homo sapiens

<400> 13
ctcgggcagc ccagtctttg ccatccttgc ccagccggtg tggtgcttgt gtgtcacagc 60

cttgtagccg ggagtcgctg ccgagtgggc gctcagtttt cgggtcgtca tggctggcta 120

cgaatacgtg agcccggagc agctggctgg ctttgataag tacaagtaca gtgctgtgga 180

taccaatcca ctttctctgt atgtcatgca tccattctgg aacactatag taaaggtatt 240

tcctacttgg ctggcgccca atctgataac tttttctggc tttctgctgg tcgtattcaa 300

ttttctgcta atggcatact ttgatcctga cttttatgcc tcagcaccag gtcacaagca 360

cgtgcctgac tgggtttgga ttgtagtggg catcctcaac ttcgtagcct acactctaga 420

tggtgtggac ggaaagcaag ctcgcagaac caattctagc actcccttag gggagctttt 480

tgatcatggc ctggatagtt ggtcatgtgt ttactttgtt gtgactgttt attccatctt 540

tggaagagga tcaactggtg tcagtgtttt tgttctttat ctcctgctat gggtagtttt 600

gttttctttc atcctgtccc actgggaaaa gtataacaca gggattcttt tcctgccatg 660

gggatatgac attagccagg tgactatttc ttttgtctac atagtgactg cagttgtggg 720
agttgaggcc tggtatgaac ctttcctgtt taatttctta tatagagacc tattcactgc 780
aatgattatt ggttgtgcat tatgtgtgac tcttccaatg agtttattaa acttttcag 840
aagctataaa aataacacct tgaaactcaa ttcagtctat gaagctatgg ttcccttatt 900
ttctccatgc ttgctgttca ttttgtctac agcgtggatc ctttggtcac cttcagatat 960
tttagagcta catcctagag tattctactt tatggttgga acagcttttg ccaacagtac 1020
atgtcagctg attgtttgcc aaatgagtag tacccggtgt ccaactttga attggttgct 1080
ggttcctctc ttcttggttg tcttagtggt aaacctagga gtagcctctt acgttgagag 1140
cattctcctg tatacattaa caactgcttt tactctggcc cacatccatt atggagtacg 1200
agtggtaaag cagctgagca gccattttca gatttacccc ttctcattga ggaaaccaaa 1260
ctcagattga ctaggaatgg aagaaaagaa tattggcctg taataatctt tctttgggca 1320
caaagaagta ctgtaaataa atgcttgtaa atatttcctc catcaccatt gaactagact 1380
gatctgcttg acagacgtgg gatctcagta tggtacttgg acagcaggaa tgatacatat 1440
aatctgaact tgggaaattt tggacctact aactctaagc ctattttatt ttttataaaa 1500
ctatgtgaca ttttggttga gcagaatgta cgttagacca gcaaaatgtt cctaatggtt 1560
ctaacttcgt gagtttacaa tgttgtgatt catgcagggt taaagatgct ttgtttttat 1620
tttttaagta ccaaaattgg tttcagaaca ctgataacac tcagaaaacc acagtgtgtt 1680
ttcatatttg gaactttgta atagcgggag tagcagtagt ccaaacctag tatagggaaa 1740
ggataaaaat aagtcacctt caccaagaga tgccaatgat taccaaacaa cagacagttg 1800
ccaaatactg gtttctcttt cccctgaaaa tggcttttgt tctcaaatga taagagagct 1860
aatacattta gctaatattc tagctctctt tattatggaa cagatcttga tagatggttt 1920
aattttctcc taaagagaaa taatcagttg agaatttgag aatgggttgt aattatcgct 1980
cacccattgg gatggttcat tgtttaaata tggcattttc ccccttcag ctgcaggttc 2040
ctgagatttg gtgcctgtga gctctgattg taggaatgca tgtgacagtc ccagtcctat 2100
ggtaatgact taggaggaat gcagataaaa gtaccttgta agataaatat aaattggagt 2160
taggaatttc atgaacctca ctatgaccaa aattaatttt ttgattcagt ttgtctgtct 2220
gtctgtcctt cccctctctt ctttttcag ggtgaggtgc tgtgtttctt atttcatacg 2280
agataaaaca gagagaagtt ctctcttctc cagcttgtcc atttccccac ttgaagaaaa 2340
cttttgatat atatgcctta ctgagtacat gccccttta atgttaacat gacttggagt 2400
aatttctgag gtttactgac aaacataaaa atccctttaa ttgtagtgta gttgttctat 2460
aaaccatatt ttttcatgat gtggatattt tcttctattt ctttgtcttc atttaatttg 2520
gtggtggtga actttacttg ctgattttct tttatttttc actgaatgaa gtttgtgctt 2580

gaatgaagag tgtatcttaa acccccttt tttggacagg ctgcacttgg ataaaatagg 2640

caccactgtg ttgatatgta attaaattca taactattta tgaatgtgac cattgttaaa 2700

tttaattata gtgtatttct ttcttgaatt tcagtctgct gctttgtaaa aaataactgt 2760

aattgatatt tcatttttaa aatttataaa ttccaatcat tgtcattcaa gttattagaa 2820

atgagtcacg gcactaactc atggttctat actcatttta agaaattgta aaggttgagg 2880

tgtatttggt gactttctct gagacacaga tgaacgttgg aaatagatga gaaatggatg 2940

cagaataatg cagatagtga tcccatgttt gttaattgcc atcatggaag tcagagaatt 3000

aaccaattaa gtgacggtaa ctatttttg tataccttgt tctttgtgat ctataaaatt 3060

aaactggata aaatgttaat tgggggtagg gaaaaagctt tgttttaaga tttctaagct 3120

aattatttta ccgttttgtt gaccttgtac acaagaatga gattaatctg tattatataa 3180

gaaaagaact gataaatatt tataaaatga caaacttgga ccacaggctt tcttgacctt 3240

gtacagtata gtaaaaatac cctttacttc tgacatttta aggtcacgtc ctatagtttg 3300

ggggcccttt tacttcaaag gtgtgtttct gccctctttt ggttgtatca aatttatttg 3360

aaaattgttt tgtgctgtac tttatgactt gtttccctat ggtggtaaag aatgatgcca 3420

tagtcagttt tattgacatt gttagatgaa agaaaagtag ctgtcaacaa ggggaaagtg 3480

aaattgagca attcactgtt tcaggatact tgaggatctt tagtgtactg ctgactcttt 3540

ttcttgtttt gtctgccatt cttgaaagtg catcttgaga gcccccattt gttgtgtaga 3600

gatatagcag ctgaagcagg acttcattcc agatagctgt gtgcatttat gtcatgccaa 3660

ttactataac atttccaaag tatgaaaggt ctttcaggct tatgcatgat attgtgggtg 3720

acttcatgcc ccagcctttc tccccagcat acagtgagct gccttttggg gagcaatgcc 3780

atgataagga gcatacgtgc atgttttccg tactttcaga atagagtgcc ttagctaggc 3840

cagaggctcg tattgtgttt ctgctgacca aagttgcata cacctgtaag aagtgtcttt 3900

ttaggttatg gtattaaaaa agtcgtttca ggcttatcag ataataattc tcctatattt 3960

gaggggggca tgggataaaa agatttaata atacctggat agagaatgct aatgctgcct 4020

tttgtgatcc tttaatatta tccatgcttt tataaagtgc agtacacatt ttgacattgc 4080

cacagtttac agtgtcattc ttccacgaat aggtaattga tagagtttat gggtcttagt 4140

cttaagggta acagagtgaa taccttccaa ataaaaagct aaatattcct ccttgtcatt 4200

ttaatgtcat tattgaaata cttttaattc taattttcaa aagaaagcta aatgctgatt 4260

gtggtttagg atgttacagt actaatacca tcatgttatt tattctaata tagaggcatt 4320

gagaggaggg gagtgtatgt ctagagttac cccaattcta aataggtacc aaaagagtat 4380

tgattatgac ctggagcctg agctttggga gaatatacaa aggaaaaagc aggcatagtt 4440

tccactttaa agggaagaag ggactttaac atcctatgaa tttgaaacat cctatgaatt 4500

tgaaagtttc ctaagtgttg ttttcaaaga actggtttat tttatagtca gtcaaaatgc 4560
atttggaatg taccttgctt ccataacaga ctttgactct atgacctatg aagaataaaa 4620
agataatctg aaggtgtttg agtcagaggc tacttttccc ctttatattt tatatagaag 4680
cctatttgta attcttcctc ttcgttcttc acctgggttt cctcttatct acccaccacc 4740
taggacagag gctgcccact caggtcttgc cttgttgccc ctagatgttg ccacattcac 4800
atctccttga tggccttaga tctcttgact agtcaaggac tgggaacgtg ccaggggaa 4860
caaagtaagc ttgatgggct ggaacttttg actttgggtc cccaaacagc caaaaacaga 4920
tacagaaatc tttggttgtg tggatgcctg tgttcagagt ggccttatct tgacctcagt 4980
tctggattta agtactctat agaaaatttg tgaagtcaga ccagaatttt ctttgtccca 5040
tgaggtgtct ttgtaagtca gcatgctctt gcttgaacac cctgtgggtg aataacttta 5100
gtaagttagc tgccagcatt agaaccttcc aggaagaatt ttaactagta tgctacattt 5160
gaagcactct gacatgtaca tttaaagtga cagttttaaa agtaaaatgg tgttaaatct 5220
ctgaaacttg aacgtatgat gctctggaca ttttaggaag cttaagtaaa tttctgattt 5280
tcctttattg attatgtaac ccacgtgatt tgcatatttt agtttttatg tttgacagtg 5340
gtacaaatga ctgacattgg caggcagtta ttgtagggta tatggttaat cacgattgtc 5400
taacttttcc agagaaagtg gtgatatgct tcttccttgc atggagctgt gggaattgga 5460
tttgttggct cttcatttcc agtgtttcta cagtgcagaa taaatgtaac tttttaagcc 5520
tgcatgctag taatatgctt tgctctttag taaacaaata tgcaatagat agtaatctat 5580
ctacttttt taaagcaggg atctgctccc tatctagaca gatcttttaa tgcattaata 5640
tttagagtag caaaagaagt tttgccattt gtagctctct cacaggtagc ttaagcgagg 5700
gggttgtgtt gcatgtgctt agaaaggcca ggatcccaga attggaaggt gccacatata 5760
cttttagaat attttaacaa aggcttgtct ctaggaaaat ccacttggac tgcagcagga 5820
ttgttctggt cttccacaga ccattcacat ggtctttctc tccagagaat cattcttaga 5880
aagccaggct aaatcagagt agtgctgttt ttaaagcaag tatattggat gatggctata 5940
gggtacctta atgctgtgtt gaaattggaa gcagactgct cttcagcagc acatatactg 6000
caactgaagt tagagacagt actgaagacg agggagtagc taaatgctgt ggatatttca 6060
tgtctgtttg gctaatgtct tctcttgagg agtttctatg ggtagcaatt aaaattagaa 6120
tttattttaa ggtttttatt aaaatatata caatttgcca taaattctag attctgatat 6180
tcttgacaag cacttttga atatctcaag agtaagttct tgacctggaa catatatagt 6240
tagtacttga tgtcactaac caagttgtag aaggagcctg ttataattct gcaagatctg 6300
tgaatagcat tataaatctg gggcaggtgc ctgtaatccc agcactttgg gaagccaagg 6360
caggtggatt gcttgagccc aggagttcga caccaacctg agcaacatgg cgaaaccccg 6420

tctgtaataa aaatacaaaa aaattagctg ggcatggtgg catgtgcctg taatcccagt 6480

tactggggag gctgaggcag gagaatcact tgaacctggg agatggagat tgcagtgagc 6540

agagattgtg ccactgcact ccagccagag tgacagagca agactccatc tcaaaaaata 6600

ataataaaaa taaacaaatc tgtttaaatt actgttgcct ttcagcatta gactttgagt 6660

ttaataacta caaattgaga ctgctcacaa tattaacttt ttttgtaggt tattttgttg 6720

tttagataac ttgcctccct agaggagcat tcaggagata aagacctagc tacatgtaat 6780

gatatgatca tttcaaaaat gtgccaagaa agcaaaattc attattgaac tctaaatttc 6840

tgggtgtttt ttttttaaa gtagcatttt ctctgggtaa agggaaaggc aatgaatgat 6900

tgccaacatg taaactccct gctgcccgcc ttcccccagt cccctccatc taacataata 6960

cagtaaattt gtagccagtt gtagaaaaag aaattgatat ctttctgagt aaggtttcat 7020

gccctgtgac taagaatagg tggaggaatc atggccaaat caaatatgaa ttgtttagct 7080

tggtcctgtt gcagttggct ttctgtagtg ttctgaaaca agaggacgat tccattcctt 7140

ctcaggaacc tagaaacaac accgcttgag caactggaat atgtttgctg caagcagaat 7200

attttgggga gaggaagagt agtttaattc aagtagttta attgaacata ttagtcattg 7260

gtctgtctgg gacgtgcagt gttcatagta gcaatgtatg taccatttat tttatctggt 7320

tgtgtgtgat gtgtgtgtat gcctatttaa tatgtacaca tatattcact acatatgtat 7380

gtatgatata ttcatatata catgcagtct gcttgattat cagcaaaatg gtcagccttt 7440

atcagatagt ttcttcatgt ggagttcatc tgcatgtggc ccttactctg aagcctcttc 7500

ctgatctgga gccacagtct gtctgtcttc cagttcatct cagtcctcga gaaaggccct 7560

ttaaatatgt cactttccca ttttccttta accatgggtt gtgtgagcca gaaagagctt 7620

tgagaaagat ggctgcttcc accagggtgg aggcttctag gtctgcatga tgatggggcc 7680

cgtttctggc cagagggtgg ctctgggagc agttgtgctg cgggcttgct gggggagaac 7740

tctaactgtt gcagaaacag agcttcatgg cttgcttaaa ttacttagct ggaatatttt 7800

aaagtgtcag ataatgtgat gtacaaagag agtatgccga tgcatttcat tgtttttgca 7860

ttatctgtat cagtgtttcc cattgctagc ccgaataatc aagaccagac tgtttcttgt 7920

cactggattc tggaagcctt gccctttcaa ctggatttat gtttgtttct gctgctgtga 7980

aaccccaaag taatgcagta ggttttaatt gaaaattgtt taattctatt attgataatt 8040

atttattgta ataaaagtaa aagagtaatt tttaaatcct ttt 8083

<210> 14
<211> 386
<212> PRT
<213> Homo sapiens

<400> 14

Met Ala Gly Tyr Glu Tyr Val Ser Pro Glu Gln Leu Ala Gly Phe Asp
1 5 10 15

Lys Tyr Lys Tyr Ser Ala Val Asp Thr Asn Pro Leu Ser Leu Tyr Val
20 25 30

Met His Pro Phe Trp Asn Thr Ile Val Lys Val Phe Pro Thr Trp Leu
35 40 45

Ala Pro Asn Leu Ile Thr Phe Ser Gly Phe Leu Leu Val Val Phe Asn
50 55 60

Phe Leu Leu Met Ala Tyr Phe Asp Pro Asp Phe Tyr Ala Ser Ala Pro
65 70 75 80

Gly His Lys His Val Pro Asp Trp Val Trp Ile Val Val Gly Ile Leu
85 90 95

Asn Phe Val Ala Tyr Thr Leu Asp Gly Val Asp Gly Lys Gln Ala Arg
100 105 110

Arg Thr Asn Ser Ser Thr Pro Leu Gly Glu Leu Phe Asp His Gly Leu
115 120 125

Asp Ser Trp Ser Cys Val Tyr Phe Val Val Thr Val Tyr Ser Ile Phe
130 135 140

Gly Arg Gly Ser Thr Gly Val Ser Val Phe Val Leu Tyr Leu Leu Leu
145 150 155 160

Trp Val Val Leu Phe Ser Phe Ile Leu Ser His Trp Glu Lys Tyr Asn
165 170 175

Thr Gly Ile Leu Phe Leu Pro Trp Gly Tyr Asp Ile Ser Gln Val Thr
180 185 190

Ile Ser Phe Val Tyr Ile Val Thr Ala Val Val Gly Val Glu Ala Trp
195 200 205

Tyr Glu Pro Phe Leu Phe Asn Phe Leu Tyr Arg Asp Leu Phe Thr Ala
210 215 220

Met Ile Ile Gly Cys Ala Leu Cys Val Thr Leu Pro Met Ser Leu Leu
225 230 235 240

Asn Phe Phe Arg Ser Tyr Lys Asn Asn Thr Leu Lys Leu Asn Ser Val
245 250 255

Tyr Glu Ala Met Val Pro Leu Phe Ser Pro Cys Leu Leu Phe Ile Leu
260 265 270

Ser Thr Ala Trp Ile Leu Trp Ser Pro Ser Asp Ile Leu Glu Leu His
275 280 285

Pro Arg Val Phe Tyr Phe Met Val Gly Thr Ala Phe Ala Asn Ser Thr
290 295 300

Cys Gln Leu Ile Val Cys Gln Met Ser Ser Thr Arg Cys Pro Thr Leu
305 310 315 320

Asn Trp Leu Leu Val Pro Leu Phe Leu Val Val Leu Val Val Asn Leu
325 330 335

Gly Val Ala Ser Tyr Val Glu Ser Ile Leu Leu Tyr Thr Leu Thr Thr
340 345 350

Ala Phe Thr Leu Ala His Ile His Tyr Gly Val Arg Val Val Lys Gln
355 360 365

Leu Ser Ser His Phe Gln Ile Tyr Pro Phe Ser Leu Arg Lys Pro Asn
370 375 380

Ser Asp
385

<210> 15
<211> 2788
<212> DNA
<213> Homo sapiens

<400> 15
atagaaggaa gaggagcgaa catggcagcg cgttggcggt tttggtgtgt ctctgtgacc 60

atggtggtgg cgctgctcat cgtttgcgac gttccctcag cctctgccaa aagaaagaag 120

gagatggtgt tatctgaaaa ggttagtcag ctgatggaat ggactaacaa aagacctgta 180

ataagaatga atggagacaa gttccgtcgc cttgtgaaag ccccaccgag aaattactcc 240

gttatcgtca tgttcactgc tctccaactg catagacagt gtgtcgtttg caagcaagct 300

gatgaagaat tccagatcct ggcaaactcc tggcgatact ccagtgcatt caccaacagg 360

atatttttg ccatggtgga ttttgatgaa ggctctgatg tatttcagat gctaaacatg 420

aattcagctc caactttcat caactttcct gcaaaaggga aacccaaacg gggtgataca 480

tatgagttac aggtgcgggg tttttcagct gagcagattg cccggtggat cgccgacaga 540

actgatgtca atattagagt gattagaccc ccaaattatg ctggtcccct tatgttggga 600

ttgcttttgg ctgttattgg tggacttgtg tatcttcgaa gaagtaatat ggaatttctc 660

tttaataaaa ctggatgggc ttttgcagct ttgtgttttg tgcttgctat gacatctggt 720
caaatgtgga accatataag aggaccacca tatgcccata agaatcccca cacgggacat 780
gtgaattata tccatggaag cagtcaagcc cagtttgtag ctgaaacaca cattgttctt 840
ctgtttaatg gtggagttac cttaggaatg gtgcttttat gtgaagctgc tacctctgac 900
atggatattg gaaagcgaaa gataatgtgt gtggctggta ttggacttgt tgtattattc 960
ttcagttgga tgctctctat ttttagatct aaatatcatg gctacccata cagctttctg 1020
atgagttaaa aaggtcccag agatatatag acactggagt actggaaatt gaaaaacgaa 1080
aatcgtgtgt gtttgaaaag aagaatgcaa cttgtatatt ttgtattacc tctttttttc 1140
aagtgattta aatagttaat catttaacca aagaagatgt gtagtgcctt aacaagcaat 1200
cctctgtcaa aatctgaggt atttgaaaat aattatcctc ttaaccttct cttcccagtg 1260
aactttatgg aacatttaat ttagtacaat taagtatatt ataaaaattg taaaactact 1320
actttgtttt agttagaaca aagctcaaaa ctactttagt taacttggtc atctgatttt 1380
atattgcctt atccaaagat ggggaaagta agtcctgacc aggtgttccc acatatgcct 1440
gttacagata actacattag gaattcattc ttagcttctt catctttgtg tggatgtgta 1500
tactttacgc atctttcctt ttgagtagag aaattatgtg tgtcatgtgg tcttctgaaa 1560
atggaacacc attcttcaga gcacacgtct agccctcagc aagacagttg tttctcctcc 1620
tccttgcata tttcctactg aaatacagtg ctgtctatga ttgtttttgt tttgttgttt 1680
ttttgagacg gtctcgctgt gtcacacagg ctggagtgca gtggcgtgag ctcggctgac 1740
tgcaaactct gcctcccagg tttaagcgat tctcctgtca cagcttccca agtagctggg 1800
atttacaggt gtgcaccgcc atgccaggct aatttttgtg tttttagtag agacagggtt 1860
tcgccaagtt gtccaggctg gtcttgaact cctgggctca agtgatccgc ccgcctcagt 1920
ctcccaaagt gcgaggatga catgtgtgag ctaccacacc agcaatgtct atgcttctcg 1980
atagctgtga acatgaaaag acatctattg ggagtccgag gcaggtggat tgcttgaggc 2040
caggagttag agaccagcct ggccaacaag gcaaaacccc gtctctacta aaaatatgaa 2100
aattagctgg gcttggtggc tcatgcctat aatcctagct acttgggagg ctgaggcacg 2160
agacttgctt aatacctggg aggcggagat tgcagtgagc cgagatcacg ctactgcgct 2220
ccagcctgag tgatagagtg agactctgtc tcaaaaaaaa gtatctctaa atacaggatt 2280
ataatttctg cttgagtatg gtgttaacta ccttgtattt agaaagattt cagattcatt 2340
ccatctcctt agttttcttt taaggtgacc catctgtgat aaaaatatag cttagtgcta 2400
aaatcagtgt aacttataca tggcctaaaa tgtttctaca aattagagtt tgtcacttat 2460
tccatttgta cctaagagaa aaataggctc agttagaaaa ggactccctg gccaggcgca 2520
gtgacttacg cctgtaatct cagcactttg ggaggccaag gcaggcagat cacgaggtca 2580

ggagttcgag accatcctgg ccaacatggt gaaaccccgt ctctactaaa aatataaaaa 2640

ttagctgggt gtggtggcag gagcctgtaa tcccagctac acaggaggct gaggcacgag 2700

aatcacttga actcaggaga tggaggtttc agtgagccaa gatcacacca ctgcactcca 2760

gcctggcaac agagcgagac tccatctc 2788

<210> 16
<211> 335
<212> PRT
<213> Homo sapiens

<400> 16

Met Ala Ala Arg Trp Arg Phe Trp Cys Val Ser Val Thr Met Val Val
1 5 10 15

Ala Leu Leu Ile Val Cys Asp Val Pro Ser Ala Ser Ala Lys Arg Lys
20 25 30

Lys Glu Met Val Leu Ser Glu Lys Val Ser Gln Leu Met Glu Trp Thr
35 40 45

Asn Lys Arg Pro Val Ile Arg Met Asn Gly Asp Lys Phe Arg Arg Leu
50 55 60

Val Lys Ala Pro Pro Arg Asn Tyr Ser Val Ile Val Met Phe Thr Ala
65 70 75 80

Leu Gln Leu His Arg Gln Cys Val Val Cys Lys Gln Ala Asp Glu Glu
85 90 95

Phe Gln Ile Leu Ala Asn Ser Trp Arg Tyr Ser Ser Ala Phe Thr Asn
100 105 110

Arg Ile Phe Phe Ala Met Val Asp Phe Asp Glu Gly Ser Asp Val Phe
115 120 125

Gln Met Leu Asn Met Asn Ser Ala Pro Thr Phe Ile Asn Phe Pro Ala
130 135 140

Lys Gly Lys Pro Lys Arg Gly Asp Thr Tyr Glu Leu Gln Val Arg Gly
145 150 155 160

Phe Ser Ala Glu Gln Ile Ala Arg Trp Ile Ala Asp Arg Thr Asp Val
165 170 175

Asn Ile Arg Val Ile Arg Pro Pro Asn Tyr Ala Gly Pro Leu Met Leu
180 185 190

Gly Leu Leu Leu Ala Val Ile Gly Gly Leu Val Tyr Leu Arg Arg Ser
195 200 205

Asn Met Glu Phe Leu Phe Asn Lys Thr Gly Trp Ala Phe Ala Ala Leu
210 215 220

Cys Phe Val Leu Ala Met Thr Ser Gly Gln Met Trp Asn His Ile Arg
225 230 235 240

Gly Pro Pro Tyr Ala His Lys Asn Pro His Thr Gly His Val Asn Tyr
245 250 255

Ile His Gly Ser Ser Gln Ala Gln Phe Val Ala Glu Thr His Ile Val
260 265 270

Leu Leu Phe Asn Gly Gly Val Thr Leu Gly Met Val Leu Leu Cys Glu
275 280 285

Ala Ala Thr Ser Asp Met Asp Ile Gly Lys Arg Lys Ile Met Cys Val
290 295 300

Ala Gly Ile Gly Leu Val Val Leu Phe Phe Ser Trp Met Leu Ser Ile
305 310 315 320

Phe Arg Ser Lys Tyr His Gly Tyr Pro Tyr Ser Phe Leu Met Ser
325 330 335

<210> 17
<211> 3239
<212> DNA
<213> Homo sapiens

<400> 17
gcttggggaa cgaggagcag gacgcggcct cggtggggcc cgggccgaac ggctgcggac 60

acctgggcgc cgaggagccg agcgccgccg tctccggcat ggatcagtgc gtgacggtgg 120

agcgcgagct ggagaaggtg ctgcacaagt tctcaggcta cgggcagctg tgcgagcgcg 180

gcctggagga gctcatcgac tacaccggcg gcctcaagca cgagatcctg cagagccacg 240

gccaagatgc tgaattatca gggacacttt cacttgtttt gacacagtgc tgtaaaagaa 300

taaaggatac tgttcaaaaa ttggcctccg accacaaaga catccacagc agtgtttctc 360

gggttggaaa agccattgat aagaattttg attctgacat tagcagtgtg ggaatagatg 420

gctgctggca ggcagacagc caaaggcttc tcaatgaagt gatggtggag cacttctttc 480

gacaaggaat gctggatgtg gctgaggagc tctgtcagga atctggtctt tctgtagacc 540

caagtcagaa ggaaccattt gtggagttaa atagaatatt agaggcatta aaggtcagag 600

ttctgagacc tgctctggag tgggcagtgt caaaccggga aatgcttata gctcaaaaca 660

gctccttgga atttaagcta cacagactgt attttattag cttgttaatg ggtggaacca 720
caaatcagcg agaggcatta caatatgcta aaaattttca gccatttgcc ctaaatcatc 780
aaaaagacat tcaggttttg atgggaagcc ttgtgtacct gagacaaggg attgagaact 840
caccatatgt tcacctactt gatgcaaacc agtgggctga tatctgtgac atctttacac 900
gggatgcttg tgccctcctg gggctctccg tggagtcccc tctcagtgtc agtttctcag 960
caggttgtgt ggcgctgcca gctttaatta acatcaaagc cgtgattgaa cagaggcagt 1020
gtactggagt ttggaaccag aaagatgaat tacctattga agtggacctt ggtaaaaagt 1080
gctggtatca ctctatattt gcctgcccca ttcttcgtca gcaaacaaca gataacaatc 1140
cacccatgaa attggtctgt ggtcatatta tatcaagaga tgccctgaat aaaatgttta 1200
atggtagcaa attaaaatgt ccctactgtc caatggaaca aagtccagga gatgccaaac 1260
agatattttt ctgaagagat aactttagtt tgcaatttgt aagtgaaact gaatcgtggg 1320
tgcatttcag aagagaacgt tccatataat gcagctaacc aaggactcct gtgtttctat 1380
aagctaatgc tccagaaact ttgccaacct gttagtgtac acacactgag gggagtgctc 1440
ccggtgaata ttatcatagg gctttattat attcttggtc ttcatttctg atcaagtaaa 1500
tacaccagca gttgtcattc aatgcaggtt tttgtactta attatatggt gatttttta 1560
ctttttaaga gcagaaacgg aaattgacct ccccgccatg tgtttaatat tcctcctgct 1620
tttacttttg tcattttctt gataatcgta agccttgaga gtgtttgtga aaaagtttta 1680
tttcctgtta tgtatacata attaaatgaa aattcttcag aaaaagtttg ataaattgaa 1740
ttgtggttat gaaactaatt tgcattttta tttgcttaag aaagaaagct gtgatagatt 1800
ccagatatgc tttttgatgt tttcctctgc tccagctcca agaagtcagc acacctgcat 1860
tttagctctg catgcagccc cagcaggctg cgtgtttaag aatttcattg tttaactggc 1920
tggtgtgaga agtcttccgt tagcatagag tggaaggagt actattgttt ggttgggttt 1980
ttgtttgttt gttttttgtt tttgctttta ttgccaagag gtgcttgttt taaaagtatg 2040
tttaataaaa tgaaattcta aagttagaag tgttcttaag ttgatattta ctctcttggt 2100
cttggtagcc ctattatatc attgcagaca cagtgttgtg catgtgtgta ttgtatatgt 2160
gcaccagcat caaacattgt gtatctggaa ggaaagcagc atgttccagt tctaaaatgc 2220
agttaccaga cccatcactt taaatcctta aagttagaag ctagcaaatt ttctgtagtg 2280
cagaacgttt attgctgttt ttgtgtttga atagtataat gtttgatgcc tctcttctgc 2340
aaagcgtatc atctcattga ctgtgaatct gtatattatt ctgatggata cagataatga 2400
tcttttctct tgtgaggtat cttcatttat gcactgtcca aaaatagcca tgtgtaagag 2460
tctttctgta tgacgaacta catggaaaag acttctgtgg acataattct gaccgaaacc 2520
catgaagtta cttcagtata agaagaacgt tacacggaaa tcaccaaata ttttgcaact 2580

ttatttcttc tgacatggag tgaacatcaa taggaatact ttcaaagaaa atgaaaacac 2640

agaagcaaag agaaatgtgg cacttcacat tttaaactac agatggactt ggtttgaggg 2700

agggggaatc acagatttgg tgctaagtta attagaaact ggcagcgttt tacagtagta 2760

caccagcctg gatgtttttt ctaaaatgtt tacctgggag agctggggtt tgtttgtgag 2820

gagaaagagt actgtggaaa acctctgctt gagtaccatg tggccaggcc tatgtggatg 2880

gctactccgt gctgtgcggc ttcaccagcg gttgggattg gcccagcttg gagtgcttgt 2940

gtggtccaac ctcagtctgg ccccatagtg acttttgccc catgattctg cttcactgtt 3000

ggaatcctct ttgaagttcc ccctctcttt gctaaagcag tgaaggaaga gaacagagac 3060

aaactctttg gactgtgaaa gagaaggtag agaattccag gcaacagtct gaccaagggt 3120

gtaaaccagt ttattatata tatatttttc ctttgaatta aaacctatga gtgactacaa 3180

aaaggattag actgagccga ataaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa 3239

<210> 18
<211> 391
<212> PRT
<213> Homo sapiens

<400> 18

Met Asp Gln Cys Val Thr Val Glu Arg Glu Leu Glu Lys Val Leu His
1 5 10 15

Lys Phe Ser Gly Tyr Gly Gln Leu Cys Glu Arg Gly Leu Glu Glu Leu
20 25 30

Ile Asp Tyr Thr Gly Gly Leu Lys His Glu Ile Leu Gln Ser His Gly
35 40 45

Gln Asp Ala Glu Leu Ser Gly Thr Leu Ser Leu Val Leu Thr Gln Cys
50 55 60

Cys Lys Arg Ile Lys Asp Thr Val Gln Lys Leu Ala Ser Asp His Lys
65 70 75 80

Asp Ile His Ser Ser Val Ser Arg Val Gly Lys Ala Ile Asp Lys Asn
85 90 95

Phe Asp Ser Asp Ile Ser Ser Val Gly Ile Asp Gly Cys Trp Gln Ala
100 105 110

Asp Ser Gln Arg Leu Leu Asn Glu Val Met Val Glu His Phe Phe Arg
115 120 125

Gln Gly Met Leu Asp Val Ala Glu Glu Leu Cys Gln Glu Ser Gly Leu

```
    130                 135                 140

Ser Val Asp Pro Ser Gln Lys Glu Pro Phe Val Glu Leu Asn Arg Ile
145                 150                 155                 160

Leu Glu Ala Leu Lys Val Arg Val Leu Arg Pro Ala Leu Glu Trp Ala
                165                 170                 175

Val Ser Asn Arg Glu Met Leu Ile Ala Gln Asn Ser Ser Leu Glu Phe
            180                 185                 190

Lys Leu His Arg Leu Tyr Phe Ile Ser Leu Leu Met Gly Gly Thr Thr
        195                 200                 205

Asn Gln Arg Glu Ala Leu Gln Tyr Ala Lys Asn Phe Gln Pro Phe Ala
    210                 215                 220

Leu Asn His Gln Lys Asp Ile Gln Val Leu Met Gly Ser Leu Val Tyr
225                 230                 235                 240

Leu Arg Gln Gly Ile Glu Asn Ser Pro Tyr Val His Leu Leu Asp Ala
                245                 250                 255

Asn Gln Trp Ala Asp Ile Cys Asp Ile Phe Thr Arg Asp Ala Cys Ala
            260                 265                 270

Leu Leu Gly Leu Ser Val Glu Ser Pro Leu Ser Val Ser Phe Ser Ala
        275                 280                 285

Gly Cys Val Ala Leu Pro Ala Leu Ile Asn Ile Lys Ala Val Ile Glu
    290                 295                 300

Gln Arg Gln Cys Thr Gly Val Trp Asn Gln Lys Asp Glu Leu Pro Ile
305                 310                 315                 320

Glu Val Asp Leu Gly Lys Lys Cys Trp Tyr His Ser Ile Phe Ala Cys
                325                 330                 335

Pro Ile Leu Arg Gln Gln Thr Thr Asp Asn Asn Pro Pro Met Lys Leu
            340                 345                 350

Val Cys Gly His Ile Ile Ser Arg Asp Ala Leu Asn Lys Met Phe Asn
        355                 360                 365

Gly Ser Lys Leu Lys Cys Pro Tyr Cys Pro Met Glu Gln Ser Pro Gly
    370                 375                 380
```

Asp Ala Lys Gln Ile Phe Phe
385 390

<210> 19
<211> 5565
<212> DNA
<213> Homo sapiens

<400> 19
gcccctcgct gttctgccct ttgtcttcca cgcctgcccc ttcaactcct cttgccctgc 60

aaaccctggg tcgccacctc cacatccctt ccacttctac ctgcctgcct tggcctgcat 120

tttatttca accaaggttg aaatggtagc cccacgcaaa atccgatggt gtgtggagga 180

actgcccagt cattgcagtc tgcacgcctt cttaagtgcc agatgcgagc aaggccgacc 240

aggaactgcc gatgggcttc acctgtcaga ctggactgcc acctgccaca cattcattga 300

actttcaata agcactgccg ctgtgcttgg cgttttgctg ggctccagaa ataaaaaatt 360

cagagaggcc aaagttcttc agggatgcag acaggtgata gaaccatcat ttattgagca 420

cttgctaagt gctaaatact cttacctgcg ttatcttgtg taatcctcac aaccacactc 480

aacgacggcc tattgaatgc tctctaggtg tctggcagtg ttccaaacag tttatgtgta 540

ttactttgtt tgatcttccc aaccaccaat gagcggatat tcaactcctt tgtttacact 600

gagaaaatct caaatggaga aagtgaagta cagcagctag ccaaaaaaat ccgagagaag 660

ttcaaccgtt acttggatgt ggtcaatcgg aacaagcaag ttgtagaagc atcctatacg 720

gctcacctaa cctctcccct aactgcaatt caagactgct gtactatccc accttccatg 780

atggaattcg atgggaactt taataccaat gtgtctagaa caattagttg tgatcgactt 840

tctactactg ttaatagccg ggccttcaat ccaggacgag acttaaattc agttcttgca 900

gacaacctga aatccaaccc tggaattaag tggcaatatt tcagttcaga agaaggaatt 960

ttcactgttt tcccagcaca caagttccgg tgtaagggca gctacgaaca ccgcagtaga 1020

tttctgtgtt aactgtggca gataccgtcc ggacttgctc actagaccag tgctataaga 1080

ccttcttgtc tccagccacc agtgagacaa aaaggaaaat gtccaccttt gttagcagcg 1140

tgaagtcttc agacagtcct acccagcacg cagtgggatt ccaaaaggca tttcagctga 1200

ttcgaagtac aaacaataac acaaagttcc aagcaaatac agacatggtc atcatttacc 1260

tgtcagctgg cattacatca aaggactctt cggaagaaga taaaaaagcg actctccaag 1320

tcatcaatga agaaaatagc tttctaaaca actctgtaat gattctcacc tatgccctca 1380

tgaacgatgg ggtgactggt ttgaaagagc tggcttttct gagggatcta gctgaacaga 1440

attcagggaa gtacggtgtg ccagaccgga cggccttgcc tgtgattaag ggcagcatga 1500

tggtgctgaa tcagttgagc aacctggaga ccacagtggg caggttctac acaaaccttc 1560

ccaaccggat gattgatgaa gccgtcttca gcctgccctt ctctgatgag atgggagatg 1620

gtttgataat gactgtgagt aaaccctgtt attttggaaa cctacttctg ggaattgtag 1680
gtgtggacgt gaatctggct tacattcttg aagacgtgac gtattaccaa gactctttgg 1740
cttcctatac ttttctcata gacgacaaag gatatacact tatgcaccca tctcttacca 1800
ggccatattt attgtcagag cccccacttc atactgacat catacattat gaaaatattc 1860
caaaatttga attagttcgg caaaatatcc taagcctccc tctgggcagc cagattatcg 1920
cagtccctgt gaactcatcc ctgtcttggc acataaacaa gctgagagaa actggaaagg 1980
aagcctacaa tgttagctat gcctggaaga tggtacaaga cacttccttt attctgtgta 2040
ttgtggtgat acaaccagaa atacctgtga aacaactgaa gaacctcaac actgttccca 2100
gcagcaagct gctgtaccac cggctggatc tccttggcca gcccagtgct tgcctccact 2160
tcaaacagct ggcaacccta gaaagtccca ccatcatgct gtctgctggc agcttttcct 2220
ccccctatga gcacctcagc cagccagaga caaagcgcat ggtagagcac tacaccgcct 2280
atctcagcga caacacccgc ctcattgcta acccgggcct caaattctct gtcagaaatg 2340
aagtaatggc taccagccac gtcacagatg aatggatgac acaaatggaa atgagtagcc 2400
tgaacactta cattgtccgc cgttacatag caacacccaa tggcgtcctc agaatttatc 2460
ctggttccct catggacaaa gcatttgatc ccactaggag acaatggtat ctccatgcag 2520
tagctaatcc agggttgatt tctttgactg gtccttactt agatgttgga ggagccggtt 2580
atgttgtgac aatcagtcac acaattcatt catccagtac acagctgtct tctgggcaca 2640
ctgtggctgt gatgggcatt gacttcacac tcagatactt ctacaaagtt ctgatggacc 2700
tattacctgt ctgtaaccaa gatggtggca acaaaataag gtgcttcata atggaggaca 2760
ggggttatct ggtggcgcac ccgactctca tcgaccccaa aggacatgca cctgtggagc 2820
agcagcacat cacccacaag gagcccctgg tagcaaatga tatcctcaac caccccaact 2880
ttgtaaagaa aaacctgtgc aacagcttca gtgacagaac ggtccagagg ttttataaat 2940
tcaacaccag ccttgcgggg gatttgacga accttgtgca tggcagccac tgttccaaat 3000
acagattagc aaggatccca ggaaccaacg cgtttgttgg cattgtcaac gaaacctgcg 3060
actctcttgc cttctgtgcc tgcagcatgg tggaccgact ctgtctcaac tgtcaccgaa 3120
tggaacaaaa tgaatgtgaa tgtccttgtg agtgccctct agaggtcaat gagtgcactg 3180
gcaacctcac caatgcagag aaccgaaacc ccagctgcga ggtccacctg gagccggtga 3240
catacacagc tattgaccct ggcctgcaag atgctcttca ccagtgtgtc aacagcaggt 3300
gcagtcagag gctggaaagt ggggactgtt ttggggtgct ggattgtgaa tggtgcatgg 3360
tggacagtga tggaaagact cacctggaca aaccctactg tgccccccag aaagaatgct 3420
tcgggggggat tgtgggagcc aaaagtccct acgttgatga catgggagca ataggtgatg 3480
aggtgatcac attaaacatg attaaaagcg cccctgtggg tcctgtggct ggagggatca 3540

tgggatgcat catggtcttg gtcctggcgg tgtatgccta ccgccaccag attcatcgcc 3600
ggagccatca gcatatgtct cctcttgctg cccaagaaat gtcagtgcgt atgtccaacc 3660
tggagaatga cagagatgaa agggacgacg acagccacga agacagaggc atcatcagca 3720
acactcggtt tatagctgcg gtcatcgaac gacatgcaca cagtccagaa agaaggcgcc 3780
gctactgggg tcgatcagga acagaaagtg atcatggtta cagcaccatg agcccacagg 3840
aggacagtga aaatcctcca tgcaacaatg acccccttgtc agccggggtc gatgtgggaa 3900
accatgatga ggacttagac ctggataccc cccctcagac tgctgcccta ctaagtcaca 3960
agttccacca ctaccggtca caccacccta cacttcatca tagccaccac ttacaggcgg 4020
ccgtcacggt acacactgtc gatgcagaat gctaacaatc tcctcacctc cacgccaaga 4080
tgagatctgg gagccacaga atgttctgga aagaaaaaga accggcttaa aacccacagc 4140
aagagacctc ccttgtgttt gtgctttgtg cagagttgtt tgagtcattt cctgcctgtc 4200
gacatggtta aaaacgagag aaacaacaac acagtcacat ttgtgaagat gtgaggctgg 4260
ttctgaaatg gaggggaaat aagcctgatg aacagacctg ccataacact aatggaaggt 4320
aacagaaggc gaacctccaa acacagagac ggaacctgca agtgaagctg agccagagga 4380
atgttccaaa gagccagaag cattcagctc tccttaactg gaagagagaa aaatctgctc 4440
acccagagac tggaatgtgg cacatgcaga tacaaatgtg tgcattgaag atttcgcttt 4500
gtttcttagc ggtacctgga taccacagtt gctgtatgga actcatgtta tgctctaaac 4560
gatgcatctc agaatttcta agtaaaggat tatttttcta ctatttattg aactttcaaa 4620
cattctcaaa ctttggggaa aaggaaagga aacacaggag aagttttcag cagttgcccc 4680
gagctgtttt gtgtgtaatg aagtggttct ttgattaagg agctctattt cttatttaac 4740
tgatatccca ctgccccact ccacaaaata ggaaaatgaa gaaatctttc tctctgactt 4800
gtttacatca tttcacggaa acacatcttt gtttgtaatg cagtattctt tctctgtgtt 4860
tgacagagat ggggaggggc agaggaattt aagaggtttt aaaagaaatg ttatgtttct 4920
tatgacttgt ttccactcct cgtacaatgc tattcttagg tttctacgaa acctaatgtt 4980
agaaccgcat cctttcagct aagggagggt tggatttatt ttccttgttt tagagactac 5040
aaatttttaa atatcccatt ttgactgaga atattgacat ataagggaag aagttttcta 5100
aattgtgaaa gtctggttct taattaaaga attttttttt taatatcacg gttaaaagct 5160
gctgccagtt agccaagaca ttatccacca aattgctttg tgatttatac agggattaat 5220
caaatctggc tactataaca tggggcattg taactttaaa gtagtgtttt aattacagtg 5280
atgtatttta gactcacatt ttgtgattca aatatgttat aaaggcattc ttgcaccatg 5340
gtaaagaatg tgtgtggtaa atctccgttt atatgtagtt ggaaaaaatt cactgaataa 5400
tgttttaatg atagggtatt atgatacaat gtaaaaaaca attggttctt cagcagtaca 5460

gaaagtaaac tatatatgtg ctatcaggaa acccccttcat actgtgtata aaattgcaat 5520

ctagtgaaat aaactgtatg caatggaaaa aaaaaaaaaa aaaca 5565

<210> 20
<211> 978
<212> PRT
<213> Homo sapiens

<400> 20

Met Ser Thr Phe Val Ser Ser Val Lys Ser Ser Asp Ser Pro Thr Gln
1 5 10 15

His Ala Val Gly Phe Gln Lys Ala Phe Gln Leu Ile Arg Ser Thr Asn
20 25 30

Asn Asn Thr Lys Phe Gln Ala Asn Thr Asp Met Val Ile Ile Tyr Leu
35 40 45

Ser Ala Gly Ile Thr Ser Lys Asp Ser Ser Glu Glu Asp Lys Lys Ala
50 55 60

Thr Leu Gln Val Ile Asn Glu Glu Asn Ser Phe Leu Asn Asn Ser Val
65 70 75 80

Met Ile Leu Thr Tyr Ala Leu Met Asn Asp Gly Val Thr Gly Leu Lys
85 90 95

Glu Leu Ala Phe Leu Arg Asp Leu Ala Glu Gln Asn Ser Gly Lys Tyr
100 105 110

Gly Val Pro Asp Arg Thr Ala Leu Pro Val Ile Lys Gly Ser Met Met
115 120 125

Val Leu Asn Gln Leu Ser Asn Leu Glu Thr Thr Val Gly Arg Phe Tyr
130 135 140

Thr Asn Leu Pro Asn Arg Met Ile Asp Glu Ala Val Phe Ser Leu Pro
145 150 155 160

Phe Ser Asp Glu Met Gly Asp Gly Leu Ile Met Thr Val Ser Lys Pro
165 170 175

Cys Tyr Phe Gly Asn Leu Leu Leu Gly Ile Val Gly Val Asp Val Asn
180 185 190

Leu Ala Tyr Ile Leu Glu Asp Val Thr Tyr Tyr Gln Asp Ser Leu Ala
195 200 205

Ser Tyr Thr Phe Leu Ile Asp Asp Lys Gly Tyr Thr Leu Met His Pro
210 215 220

Ser Leu Thr Arg Pro Tyr Leu Leu Ser Glu Pro Pro Leu His Thr Asp
225 230 235 240

Ile Ile His Tyr Glu Asn Ile Pro Lys Phe Glu Leu Val Arg Gln Asn
245 250 255

Ile Leu Ser Leu Pro Leu Gly Ser Gln Ile Ile Ala Val Pro Val Asn
260 265 270

Ser Ser Leu Ser Trp His Ile Asn Lys Leu Arg Glu Thr Gly Lys Glu
275 280 285

Ala Tyr Asn Val Ser Tyr Ala Trp Lys Met Val Gln Asp Thr Ser Phe
290 295 300

Ile Leu Cys Ile Val Val Ile Gln Pro Glu Ile Pro Val Lys Gln Leu
305 310 315 320

Lys Asn Leu Asn Thr Val Pro Ser Ser Lys Leu Leu Tyr His Arg Leu
325 330 335

Asp Leu Leu Gly Gln Pro Ser Ala Cys Leu His Phe Lys Gln Leu Ala
340 345 350

Thr Leu Glu Ser Pro Thr Ile Met Leu Ser Ala Gly Ser Phe Ser Ser
355 360 365

Pro Tyr Glu His Leu Ser Gln Pro Glu Thr Lys Arg Met Val Glu His
370 375 380

Tyr Thr Ala Tyr Leu Ser Asp Asn Thr Arg Leu Ile Ala Asn Pro Gly
385 390 395 400

Leu Lys Phe Ser Val Arg Asn Glu Val Met Ala Thr Ser His Val Thr
405 410 415

Asp Glu Trp Met Thr Gln Met Glu Met Ser Ser Leu Asn Thr Tyr Ile
420 425 430

Val Arg Arg Tyr Ile Ala Thr Pro Asn Gly Val Leu Arg Ile Tyr Pro
435 440 445

Gly Ser Leu Met Asp Lys Ala Phe Asp Pro Thr Arg Arg Gln Trp Tyr
450 455 460

Leu His Ala Val Ala Asn Pro Gly Leu Ile Ser Leu Thr Gly Pro Tyr
465 470 475 480

Leu Asp Val Gly Gly Ala Gly Tyr Val Val Thr Ile Ser His Thr Ile
485 490 495

His Ser Ser Ser Thr Gln Leu Ser Ser Gly His Thr Val Ala Val Met
500 505 510

Gly Ile Asp Phe Thr Leu Arg Tyr Phe Tyr Lys Val Leu Met Asp Leu
515 520 525

Leu Pro Val Cys Asn Gln Asp Gly Gly Asn Lys Ile Arg Cys Phe Ile
530 535 540

Met Glu Asp Arg Gly Tyr Leu Val Ala His Pro Thr Leu Ile Asp Pro
545 550 555 560

Lys Gly His Ala Pro Val Glu Gln Gln His Ile Thr His Lys Glu Pro
565 570 575

Leu Val Ala Asn Asp Ile Leu Asn His Pro Asn Phe Val Lys Lys Asn
580 585 590

Leu Cys Asn Ser Phe Ser Asp Arg Thr Val Gln Arg Phe Tyr Lys Phe
595 600 605

Asn Thr Ser Leu Ala Gly Asp Leu Thr Asn Leu Val His Gly Ser His
610 615 620

Cys Ser Lys Tyr Arg Leu Ala Arg Ile Pro Gly Thr Asn Ala Phe Val
625 630 635 640

Gly Ile Val Asn Glu Thr Cys Asp Ser Leu Ala Phe Cys Ala Cys Ser
645 650 655

Met Val Asp Arg Leu Cys Leu Asn Cys His Arg Met Glu Gln Asn Glu
660 665 670

Cys Glu Cys Pro Cys Glu Cys Pro Leu Glu Val Asn Glu Cys Thr Gly
675 680 685

Asn Leu Thr Asn Ala Glu Asn Arg Asn Pro Ser Cys Glu Val His Leu
690 695 700

Glu Pro Val Thr Tyr Thr Ala Ile Asp Pro Gly Leu Gln Asp Ala Leu
705 710 715 720

```
His Gln Cys Val Asn Ser Arg Cys Ser Gln Arg Leu Glu Ser Gly Asp
                725                 730                 735

Cys Phe Gly Val Leu Asp Cys Glu Trp Cys Met Val Asp Ser Asp Gly
            740                 745                 750

Lys Thr His Leu Asp Lys Pro Tyr Cys Ala Pro Gln Lys Glu Cys Phe
        755                 760                 765

Gly Gly Ile Val Gly Ala Lys Ser Pro Tyr Val Asp Asp Met Gly Ala
    770                 775                 780

Ile Gly Asp Glu Val Ile Thr Leu Asn Met Ile Lys Ser Ala Pro Val
785                 790                 795                 800

Gly Pro Val Ala Gly Gly Ile Met Gly Cys Ile Met Val Leu Val Leu
                805                 810                 815

Ala Val Tyr Ala Tyr Arg His Gln Ile His Arg Arg Ser His Gln His
            820                 825                 830

Met Ser Pro Leu Ala Ala Gln Glu Met Ser Val Arg Met Ser Asn Leu
        835                 840                 845

Glu Asn Asp Arg Asp Glu Arg Asp Asp Asp Ser His Glu Asp Arg Gly
    850                 855                 860

Ile Ile Ser Asn Thr Arg Phe Ile Ala Ala Val Ile Glu Arg His Ala
865                 870                 875                 880

His Ser Pro Glu Arg Arg Arg Arg Tyr Trp Gly Arg Ser Gly Thr Glu
                885                 890                 895

Ser Asp His Gly Tyr Ser Thr Met Ser Pro Gln Glu Asp Ser Glu Asn
            900                 905                 910

Pro Pro Cys Asn Asn Asp Pro Leu Ser Ala Gly Val Asp Val Gly Asn
        915                 920                 925

His Asp Glu Asp Leu Asp Leu Asp Thr Pro Pro Gln Thr Ala Ala Leu
    930                 935                 940

Leu Ser His Lys Phe His His Tyr Arg Ser His His Pro Thr Leu His
945                 950                 955                 960

His Ser His His Leu Gln Ala Ala Val Thr Val His Thr Val Asp Ala
                965                 970                 975
```

Glu Cys

<210> 21
<211> 3352
<212> DNA
<213> Homo sapiens

<400> 21
agcgggcagg gaccagatcc gcgagcccgt cagcctgcgc catgggctgc gacggccgcg 60

tgtcggggct gctccgccgc aacctgcagc ccacgctcac ctactggagc gtcttcttca 120

gcttcggcct gtgcatcgcc ttcctggggc ccacgctgct ggacctgcgc tgtcagacgc 180

acagctcgct gccccagatc tcctgggtct tcttctcgca gcagctctgc ctcctgctgg 240

gcagcgccct cgggggcgtc ttcaaaagga ccctggccca gtcactatgg gccctgttca 300

cctcctctct ggccatctcc ctggtgtttg ccgtcatccc cttctgccgc gacgtgaagg 360

tgctggcctc agtcatggcg ctggcgggct tggccatggg ctgcatcgac accgtggcca 420

acatgcagct ggtaaggatg taccagaagg actcggccgt cttcctccag gtgctccatt 480

tcttcgtggg ctttggtgct ctgctgagcc cccttattgc tgacctttc ctgtctgagg 540

ccaactgctt gcctgccaat agcacggcca acaccacctc ccgaggccac ctgttccatg 600

tctccagggt gctgggccag caccacgtag atgccaagcc ttggtccaac cagacgttcc 660

cagggctgac tccaaaggac ggggcaggga cccgagtgtc ctatgccttc tggatcatgg 720

ccatcatcaa tcttccagtg cccatggctg tgctgatgct gctgtccaag gagcggctgc 780

tgacctgctg tccccagagg aggcccctgc ttctgtctgc tgatgagctt gccttggaga 840

cacagcctcc tgagaaggaa gatgcctcct cactgccccc aaagtttcag tcacacctag 900

ggcatgagga cctgttcagc tgctgccaaa ggaagaacct cagaggagcc ccttattcct 960

tctttgccat ccacatcacg gccgccctgg tcctgttcat gacggatggg ttgacgggtg 1020

cctattccgc cttcgtgtac agctatgctg tggagaagcc cctgtctgtg ggacacaagg 1080

tggctggcta cctccccagc ctcttctggg gcttcatcac actgggccgg ctcctctcca 1140

ttcccatatc ctcaagaatg aagccggcca ccatggtttt catcaacgtg gttggcgtgg 1200

tggtgacgtt cctggtgctg cttattttct cctacaacgt cgtcttcctg ttcgtgggga 1260

cggcaagcct gggcctgttt ctcagcagca cctttcccag catgctggcc tacacggagg 1320

actcgctgca gtacaaaggc tgtgcaacca cagtgctggt gacaggggca ggagttggcg 1380

agatggtgct gcagatgctg gttggttcga tattccaggc tcagggcagc tatagtttcc 1440

tggtctgtgg cgtgatcttt ggttgtctgg cttttacctt ctatatcttg ctcctgtttt 1500

tccacaggat gcaccctgga ctcccatcag ttcctaccca agacagatca attggaatgg 1560

aaaactctga gtgctaccag aggtaaaact gggtgaagaa ggcaagagaa gactttcagc 1620

ctcttgatca ccagcacgac catactgttt cagaaagctg ggtggtggtg gaggcgctat 1680
ctcaatggct attcaagtct tctccactaa aacttggttg ggtagaggaa attaaattga 1740
gtcctggtac ctggtcaaaa tcattagaag tttacctggc ttctcaagtt atcttcttcc 1800
ctggttcaga ctgttggtaa gagctgtcca gatacccaga tgggaaggaa ggagacagcc 1860
gcgcgcttca ctccatttgt cacctcatgc atggaccata ctctgggttt gagatcattc 1920
ttcattgaag tttgtaaaaa taggttgaaa ttgtaaagct ccatgatcat tgctatgtag 1980
atatatttca atttaagcaa aacaagctgc aagttattcc ctggcatgct caaaggattt 2040
tcgtgctttt cacttaatag tccaaagtct cttaaattcc tgctgcagac atcaatagct 2100
tatctatatt ctcaaacacc aaaaggaaaa gttgaatctt gctctctttg gtatactaat 2160
gtagtggtat gctaagctgg ctcataccaa cttagaaaag ctgattgtaa aattttcatt 2220
ttgacagctg gttattaaat gcagccatta ttaaaaatca aatcatacaa acttataatt 2280
aaatcaatta catttaaaac aaaggtaata aatattcaaa gcatatcact tcctaatttg 2340
atcttgatgc tcttgaggta atttacgtcc atggtacctg tgtggtggaa ttactatata 2400
tgatggtgtg ctactgtgca ccttgtctca actccactct tcgtgatagc atgttggtag 2460
cttgaaatca gcctggtggg agtattacca tggacactgg caaaagctac agatcccgga 2520
gagccagtgg ttaaacattt accagcatac cactgctagt aatcaaggct aactggtcca 2580
gaaatcgccc aggagaatga aatggatgtt ccatttttt ctactgacat tgactagcat 2640
ataaaaggta tagaaacagc actaagactt tctgaaaata cctaatgaaa attttacatc 2700
ttttttgtt tttgtttttg tttttgagac agagtctcac tctgttgcca ggctggaatg 2760
cagtggcgtc atctcggctc actgcaacct ccgcctccca gtttcaagcg attctcctgc 2820
ctcagcctcc caagtagctg tgactacagg cgcatgccat cacgcacagc taatttttg 2880
tatttttagt agagacggga tttcaccgtg ttagccagga tggtcttgat ctcctgacct 2940
catgatctgc ccgtctcggc ctcccaaagt gctgggatta caggcgtgag ccactgcacc 3000
cagccaaaaa ttttacatct tttatagagg gaaaaaaact ctttatacca tggcaaggcc 3060
ttttctttca caaaaagctg ggcctactga acaattcaag ctgtgcagta gtagactgaa 3120
agcaggattt gttgaggagt tacagctcct gtccagagca aatcctgtag tgatacaagg 3180
agaatgtaaa cttgccagct tagacaggga tcagtcctga gactgctggc agtagcaaat 3240
ggctattaga gtaactgtat aatggttttg cctgcacttt ctctatgtat atacaaatgt 3300
acatgtataa atataaaaat taagtgatca tgaaagaaaa aaaaaaaaaa aa 3352

<210> 22
<211> 403
<212> PRT
<213> Homo sapiens

<400> 22

Met Ala Leu Ala Gly Leu Ala Met Gly Cys Ile Asp Thr Val Ala Asn
1 5 10 15

Met Gln Leu Val Arg Met Tyr Gln Lys Asp Ser Ala Val Phe Leu Gln
20 25 30

Val Leu His Phe Phe Val Gly Phe Gly Ala Leu Leu Ser Pro Leu Ile
35 40 45

Ala Asp Pro Phe Leu Ser Glu Ala Asn Cys Leu Pro Ala Asn Ser Thr
50 55 60

Ala Asn Thr Thr Ser Arg Gly His Leu Phe His Val Ser Arg Val Leu
65 70 75 80

Gly Gln His His Val Asp Ala Lys Pro Trp Ser Asn Gln Thr Phe Pro
85 90 95

Gly Leu Thr Pro Lys Asp Gly Ala Gly Thr Arg Val Ser Tyr Ala Phe
100 105 110

Trp Ile Met Ala Ile Ile Asn Leu Pro Val Pro Met Ala Val Leu Met
115 120 125

Leu Leu Ser Lys Glu Arg Leu Leu Thr Cys Cys Pro Gln Arg Arg Pro
130 135 140

Leu Leu Leu Ser Ala Asp Glu Leu Ala Leu Glu Thr Gln Pro Pro Glu
145 150 155 160

Lys Glu Asp Ala Ser Ser Leu Pro Pro Lys Phe Gln Ser His Leu Gly
165 170 175

His Glu Asp Leu Phe Ser Cys Cys Gln Arg Lys Asn Leu Arg Gly Ala
180 185 190

Pro Tyr Ser Phe Phe Ala Ile His Ile Thr Ala Ala Leu Val Leu Phe
195 200 205

Met Thr Asp Gly Leu Thr Gly Ala Tyr Ser Ala Phe Val Tyr Ser Tyr
210 215 220

Ala Val Glu Lys Pro Leu Ser Val Gly His Lys Val Ala Gly Tyr Leu
225 230 235 240

Pro Ser Leu Phe Trp Gly Phe Ile Thr Leu Gly Arg Leu Leu Ser Ile

245 250 255

Pro Ile Ser Ser Arg Met Lys Pro Ala Thr Met Val Phe Ile Asn Val
260 265 270

Val Gly Val Val Val Thr Phe Leu Val Leu Leu Ile Phe Ser Tyr Asn
275 280 285

Val Val Phe Leu Phe Val Gly Thr Ala Ser Leu Gly Leu Phe Leu Ser
290 295 300

Ser Thr Phe Pro Ser Met Leu Ala Tyr Thr Glu Asp Ser Leu Gln Tyr
305 310 315 320

Lys Gly Cys Ala Thr Thr Val Leu Val Thr Gly Ala Gly Val Gly Glu
325 330 335

Met Val Leu Gln Met Leu Val Gly Ser Ile Phe Gln Ala Gln Gly Ser
340 345 350

Tyr Ser Phe Leu Val Cys Gly Val Ile Phe Gly Cys Leu Ala Phe Thr
355 360 365

Phe Tyr Ile Leu Leu Leu Phe Phe His Arg Met His Pro Gly Leu Pro
370 375 380

Ser Val Pro Thr Gln Asp Arg Ser Ile Gly Met Glu Asn Ser Glu Cys
385 390 395 400

Tyr Gln Arg

<210> 23
<211> 1893
<212> DNA
<213> Homo sapiens

<400> 23
gtgctgttcc tgttctgtgt ggccctatac caggtgcact atgaatggtg gaggaaaatt 60

ctaaaatatt tttggatgtc agtggttatt tacactatgc tggtgcttat ctttatatac 120

acatatcagt ttgaggactt cccaggcctg tggcaaaata tgactggatt gaaaaaagaa 180

aagcttgagg atcttggctt aaagcagttt actgtggctg aactattcac tcgcatattc 240

atcccaacct cctttctgct ggtgtgcatt ttacacctgc actacttcca tgaccggttc 300

cttgaactca cagacctcaa gtccattccc agcaaagaag acaacaccat ctacagccat 360

gccaaagtca atggtcgcgt atatctaata ataaatagca tcaagaaaaa attaccaatc 420

caccaaaatg aactggccca cccggaagga agcctcccgg acctcaccat gatgcatctg 480

actgccagcc tggagaagcc ggaggtgagg aagttggctg agcctgggga ggagaagctt 540

gagggctact ctgaaaaagc ccagaagggt gatcttggga aagacagcga ggagtcagag 600

gaggacggag aggaagagga ggaatccgag gaggaggaag aaacatcaga cttaaggaac 660

aaatggcacc tggtgattga ccgcctcact gtgctcttct taaaattcct ggagtatttt 720

cacaagctgc aggtgttcat gtggtggatt ttggagttgc acatcatcaa aatcgtttcc 780

tcttacatta tctgggtttc tgtgaaagag gtgtctctgt tcaactatgt atttttgatt 840

tcttgggctt ttgctctgcc gtacgccaag ctgcgccgtc tggcttcaag tgtctgcacg 900

gtctggacgt gtgtgatcat cgtctgcaaa atgtcgtacc agcttcaaac cattaagcct 960

gagaacttct ctgttaactg ttccttgcca aatgaaaatc aaacaaacat ccccttaat 1020

gagttgaaca agtctctgct ctacagcgct cctatcgatc ctacagagtg ggtcggcctg 1080

cggaagtctt cgcctctgct agtctacctg aggaataacc tcctgatgct ggctatcctg 1140

gcctttgaag tcaccattta ccgccatcag gaatactatc gaggtcgaaa taacctgacg 1200

gcccctgtgt ctagaactat ctttcatgac attacaagac tacatctaga tgatggactt 1260

attaattgtg ccaaatattt cattaattac ttcttttaca agtttggtct ggagacctgt 1320

ttcctaatgt cagttaacgt cattggccag cgaatggatt tctatgccat gatccacgcc 1380

tgctggctga tcgctgtctt atatagacgc agaaggaaag ccatcgcaga gatctggccc 1440

aagtactgct gcttcctggc atgcatcatc accttccagt atttcatctg cattggcatc 1500

ccacctgctc cttgccgaga ttacccgtgg agattcaagg gtgccagctt caatgacaac 1560

atcataaagt ggctgtactt cccagatttc attgtgcggc ccaaccctgt gtttctcgtc 1620

tatgacttca tgctgcttct gtgtgcctcc ttacaacggc agatttttga ggatgagaac 1680

aaggctgcag tgcgaatcat ggcaggtgac aatgtcgaga tctgcatgaa tcttgatgcg 1740

gcctccttca gccaacataa ccctggtgcc accactgcaa cttttgtatt cacaggcatc 1800

ccatctccat cacagataaa atcttaggag ataaacacat tcaaaaagga atgagataaa 1860

aagaataagg caataaatgt tgattggaac ctc 1893

<210> 24
<211> 583
<212> PRT
<213> Homo sapiens

<400> 24

Met Ser Val Val Ile Tyr Thr Met Leu Val Leu Ile Phe Ile Tyr Thr
1 5 10 15

Tyr Gln Phe Glu Asp Phe Pro Gly Leu Trp Gln Asn Met Thr Gly Leu

```
          20                  25                  30

Lys Lys Glu Lys Leu Glu Asp Leu Gly Leu Lys Gln Phe Thr Val Ala
        35                  40                  45

Glu Leu Phe Thr Arg Ile Phe Ile Pro Thr Ser Phe Leu Leu Val Cys
    50                  55                  60

Ile Leu His Leu His Tyr Phe His Asp Arg Phe Leu Glu Leu Thr Asp
65                  70                  75                  80

Leu Lys Ser Ile Pro Ser Lys Glu Asp Asn Thr Ile Tyr Ser His Ala
                85                  90                  95

Lys Val Asn Gly Arg Val Tyr Leu Ile Ile Asn Ser Ile Lys Lys Lys
            100                 105                 110

Leu Pro Ile His Gln Asn Glu Leu Ala His Pro Glu Gly Ser Leu Pro
        115                 120                 125

Asp Leu Thr Met Met His Leu Thr Ala Ser Leu Glu Lys Pro Glu Val
    130                 135                 140

Arg Lys Leu Ala Glu Pro Gly Glu Glu Lys Leu Glu Gly Tyr Ser Glu
145                 150                 155                 160

Lys Ala Gln Lys Gly Asp Leu Gly Lys Asp Ser Glu Glu Ser Glu Glu
                165                 170                 175

Asp Gly Glu Glu Glu Glu Glu Ser Glu Glu Glu Glu Glu Thr Ser Asp
            180                 185                 190

Leu Arg Asn Lys Trp His Leu Val Ile Asp Arg Leu Thr Val Leu Phe
        195                 200                 205

Leu Lys Phe Leu Glu Tyr Phe His Lys Leu Gln Val Phe Met Trp Trp
    210                 215                 220

Ile Leu Glu Leu His Ile Ile Lys Ile Val Ser Ser Tyr Ile Ile Trp
225                 230                 235                 240

Val Ser Val Lys Glu Val Ser Leu Phe Asn Tyr Val Phe Leu Ile Ser
                245                 250                 255

Trp Ala Phe Ala Leu Pro Tyr Ala Lys Leu Arg Arg Leu Ala Ser Ser
            260                 265                 270
```

Val Cys Thr Val Trp Thr Cys Val Ile Ile Val Cys Lys Met Ser Tyr
275 280 285

Gln Leu Gln Thr Ile Lys Pro Glu Asn Phe Ser Val Asn Cys Ser Leu
290 295 300

Pro Asn Glu Asn Gln Thr Asn Ile Pro Phe Asn Glu Leu Asn Lys Ser
305 310 315 320

Leu Leu Tyr Ser Ala Pro Ile Asp Pro Thr Glu Trp Val Gly Leu Arg
325 330 335

Lys Ser Ser Pro Leu Leu Val Tyr Leu Arg Asn Asn Leu Leu Met Leu
340 345 350

Ala Ile Leu Ala Phe Glu Val Thr Ile Tyr Arg His Gln Glu Tyr Tyr
355 360 365

Arg Gly Arg Asn Asn Leu Thr Ala Pro Val Ser Arg Thr Ile Phe His
370 375 380

Asp Ile Thr Arg Leu His Leu Asp Asp Gly Leu Ile Asn Cys Ala Lys
385 390 395 400

Tyr Phe Ile Asn Tyr Phe Phe Tyr Lys Phe Gly Leu Glu Thr Cys Phe
405 410 415

Leu Met Ser Val Asn Val Ile Gly Gln Arg Met Asp Phe Tyr Ala Met
420 425 430

Ile His Ala Cys Trp Leu Ile Ala Val Leu Tyr Arg Arg Arg Arg Lys
435 440 445

Ala Ile Ala Glu Ile Trp Pro Lys Tyr Cys Cys Phe Leu Ala Cys Ile
450 455 460

Ile Thr Phe Gln Tyr Phe Ile Cys Ile Gly Ile Pro Pro Ala Pro Cys
465 470 475 480

Arg Asp Tyr Pro Trp Arg Phe Lys Gly Ala Ser Phe Asn Asp Asn Ile
485 490 495

Ile Lys Trp Leu Tyr Phe Pro Asp Phe Ile Val Arg Pro Asn Pro Val
500 505 510

Phe Leu Val Tyr Asp Phe Met Leu Leu Leu Cys Ala Ser Leu Gln Arg
515 520 525

Gln Ile Phe Glu Asp Glu Asn Lys Ala Ala Val Arg Ile Met Ala Gly
530 535 540

Asp Asn Val Glu Ile Cys Met Asn Leu Asp Ala Ala Ser Phe Ser Gln
545 550 555 560

His Asn Pro Gly Ala Thr Thr Ala Thr Phe Val Phe Thr Gly Ile Pro
565 570 575

Ser Pro Ser Gln Ile Lys Ser
580

<210> 25
<211> 2763
<212> DNA
<213> Homo sapiens

<400> 25
agccgggcag ctgcagcgga gccgcggagc gggcggcggg gcccaggctg tgcgcttggg 60

gagcgcggaa tgtgaggctt ggcgggccgc agcacgctcg gacgggccag gggcggcgac 120

ccctcgcgga cgcccggctg cgcgccgggc cggggacttg cccttgcacg ctccctgcgc 180

cctccagctc gccggcggga ccatgaagaa gttctctcgg atgcccaagt cggagggcgg 240

cagcggcggc ggagcggcgg gtggcggggc tggcggggcc ggggccgggg ccggctgcgg 300

ctccggcggc tcgtccgtgg gggtccgggt gttcgcggtc ggccgccacc aggtcaccct 360

ggaagagtcg ctggccgaag gtggattctc cacagttttc ctcgtgcgta ctcacggtgg 420

aatccgatgt gcattgaagc gaatgtatgt caataacatg ccagacctca atgtttgtaa 480

aagggaaatt acaattatga aagagctatc tggtcacaaa aatattgtgg gctatttgga 540

ctgtgctgtt aattcaatta gtgataatgt atgggaagtc cttatcttaa tggaatattg 600

tcgagctgga caggtagtga atcaaatgaa taagaagcta cagacgggtt ttacagaacc 660

agaagtgtta cagatattct gtgatacctg tgaagctgtt gcaaggttgc atcagtgtaa 720

gactccaata attcaccggg atctgaaggt agaaaatatt ttgttgaatg atggtgggaa 780

ctatgtactt tgtgactttg gcagtgccac taataaattt cttaatcctc aaaaagatgg 840

agttaatgta gtagaagaag aaattaaaaa gtatacaact ctgtcataca gagcccctga 900

aatgatcaac ctttatggag ggaaacccat caccaccaag gctgatatct gggcactggg 960

atgtctactc tataaacttt gtttcttcac tcttcctttt ggtgagagtc aggttgctat 1020

ctgtgatggc aacttcacca tcccagacaa ttctcgttac tcccgtaaca tacattgctt 1080

aataaggttc atgcttgaac cagatccgga acatagacct gatatatttc aagtgtcata 1140

ttttgcattt aaatttgcca aaaaggattg tccagtctcc aacatcaata agtgttgtaa 1200

acaattactg agacacggag cactgttaac tgaaattctt ctattccttc agctcttcct 1260

gaaccgatga ctgctagtga agcagctgct aggaaaagcc aaataaaagc cagaataaca 1320

gataccattg gaccaacaga aacctcaatt gcaccaagac aaagaccaaa ggccaactct 1380

gctactactg ccactcccag tgtgctgacc attcaaagtt cagcaacacc tgttaaagtc 1440

cttgctcctg gtgaattcgg taaccataga ccaaaagggg cactaagacc tggaaatggc 1500

cctgaaattt tattgggtca gggacctcct cagcagccgc cacagcagca tagagtactc 1560

cagcaactac agcagggaga ttggagatta cagcaactcc atttacagca tcgtcatcct 1620

caccagcagc agcagcagca gcagcagcaa cagcaacagc agcagcagca acagcaacag 1680

cagcagcagc agcagcagca gcagcaccac caccaccacc accaccacct acttcaagat 1740

gcttatatgc agcagtatca acatgcaaca cagcagcaac agatgcttca acaacaattt 1800

ttaatgcatt cggtatatca accacaacct tctgcatcac agtatcctac aatgatgccg 1860

cagtatcagc aggctttctt tcaacagcag atgctagctc aacatcagcc gtctcaacaa 1920

caggcatcac ctgaatatct tacctcccct caagagttct caccagcctt agtttcctac 1980

acttcatcac ttccagctca ggttggaacc ataatggact cctcctatag tgccaatagg 2040

tcagttgctg ataaagaggc cattgcaaat ttcacaaatc agaagaacat cagcaatcca 2100

cctgatatgt cagggtggaa tccttttgga gaggataatt tctctaagtt aacagaagag 2160

gaactattgg acagagaatt tgaccttcta agatcaagta agggacactt gaaggcttat 2220

tttgcttcac agtaaaataa cagctctatt attattcagc aaggccaaag acttttgaga 2280

atgtgtatgg aaaattcttt ttgtgcattt gagggcaaaa ttcaggccat cttcttatac 2340

atatactatc aaattatgtt gtgtgcatta gaaatcagtt gcttgatagt agctattaaa 2400

cccaatattg ctgatagtat gctaatatcc taaaacttaa atattgcata tctatgaatg 2460

ttaaattcag aatatctcta aaacatggaa aattgatgtt tcaataaaaa gggagacatt 2520

ttattatttt gccttactaa tgattttgca gctctgtttt tctgcacact caatagaaaa 2580

tattgtgggt ctgagatgcc ctttgaaaat gcctgaaaga aaacatgggc tactacatta 2640

tgttaatgtt ttgtaatgtc cttttaatga gtggtgacat aaaagggctg ctttgttgtc 2700

ctgatatggc aaaaattaaa taaattcttt ttcactttaa aaaaaaaaaa aaaaaaaaaa 2760

aaa 2763

<210> 26
<211> 355
<212> PRT
<213> Homo sapiens

<400> 26

Met Lys Lys Phe Ser Arg Met Pro Lys Ser Glu Gly Gly Ser Gly Gly
1 5 10 15

```
Gly Ala Ala Gly Gly Gly Ala Gly Gly Ala Gly Ala Gly Ala Gly Cys
            20                  25                  30

Gly Ser Gly Gly Ser Ser Val Gly Val Arg Val Phe Ala Val Gly Arg
        35                  40                  45

His Gln Val Thr Leu Glu Glu Ser Leu Ala Glu Gly Gly Phe Ser Thr
    50                  55                  60

Val Phe Leu Val Arg Thr His Gly Gly Ile Arg Cys Ala Leu Lys Arg
65                  70                  75                  80

Met Tyr Val Asn Asn Met Pro Asp Leu Asn Val Cys Lys Arg Glu Ile
                85                  90                  95

Thr Ile Met Lys Glu Leu Ser Gly His Lys Asn Ile Val Gly Tyr Leu
            100                 105                 110

Asp Cys Ala Val Asn Ser Ile Ser Asp Asn Val Trp Glu Val Leu Ile
        115                 120                 125

Leu Met Glu Tyr Cys Arg Ala Gly Gln Val Val Asn Gln Met Asn Lys
    130                 135                 140

Lys Leu Gln Thr Gly Phe Thr Glu Pro Glu Val Leu Gln Ile Phe Cys
145                 150                 155                 160

Asp Thr Cys Glu Ala Val Ala Arg Leu His Gln Cys Lys Thr Pro Ile
                165                 170                 175

Ile His Arg Asp Leu Lys Val Glu Asn Ile Leu Leu Asn Asp Gly Gly
            180                 185                 190

Asn Tyr Val Leu Cys Asp Phe Gly Ser Ala Thr Asn Lys Phe Leu Asn
        195                 200                 205

Pro Gln Lys Asp Gly Val Asn Val Val Glu Glu Glu Ile Lys Lys Tyr
    210                 215                 220

Thr Thr Leu Ser Tyr Arg Ala Pro Glu Met Ile Asn Leu Tyr Gly Gly
225                 230                 235                 240

Lys Pro Ile Thr Thr Lys Ala Asp Ile Trp Ala Leu Gly Cys Leu Leu
                245                 250                 255

Tyr Lys Leu Cys Phe Phe Thr Leu Pro Phe Gly Glu Ser Gln Val Ala
            260                 265                 270
```

Ile Cys Asp Gly Asn Phe Thr Ile Pro Asp Asn Ser Arg Tyr Ser Arg
275 280 285

Asn Ile His Cys Leu Ile Arg Phe Met Leu Glu Pro Asp Pro Glu His
290 295 300

Arg Pro Asp Ile Phe Gln Val Ser Tyr Phe Ala Phe Lys Phe Ala Lys
305 310 315 320

Lys Asp Cys Pro Val Ser Asn Ile Asn Lys Cys Cys Lys Gln Leu Leu
325 330 335

Arg His Gly Ala Leu Leu Thr Glu Ile Leu Leu Phe Leu Gln Leu Phe
340 345 350

Leu Asn Arg
355

<210> 27
<211> 4894
<212> DNA
<213> Homo sapiens

<400> 27
cgaaaacgga gaaaccccgg gtccggcgag aggggctgtg acagtcggag tcccaagctg 60

cggttcggct gctgccgaga actgcaaggt gtggaatatt tctggcttct agtccaatgc 120

caagtgtgtg acctgtggct acatgattcc ctgaaagata agaacaatgt tatgttgggg 180

atattggtct ctgggccaac ctggtatcag caccaacctg cagggaattg tggctgagcc 240

ccaggtgtgt gggttcatat ctgacagaag tgtcaaggaa gtggcctgtg ggggaaacca 300

ctctgtgttc ctgctggaag atggggaagt ttacacatgt ggtttgaaca ccaaggggca 360

actgggccat gagagggaag gaaacaagcc agaacaaatt ggagctctgg cagatcagca 420

tatcattcat gtggcatgtg gcgagtccca cagtctggcc ctcagtgacc gaggccagct 480

gttttcttgg ggtgcaggga gtgatggtca gctaggactc atgactactg aggattctgt 540

ggcagtgccc aggttaatac aaaagctgaa ccagcaaaca atattacaag tttcctgtgg 600

caactggcat tgcttggctc ttgcggctga tggccagttc ttcacctggg gaaagaacag 660

ccatgggcag cttggcttag ggaaggagtt cccctcccaa gccagcccac agagggtgag 720

gtccctggag gggatcccac tggctcaggt ggctgccgga ggggctcaca gctttgccct 780

gtctctctca ggagctgttt ttggctgggg gatgaataat gccgggcagc tagggctcag 840

tgatgaaaaa gatcgagaat ctccatgcca tgtaaaactc ttacgcacgc aaaaagttgt 900

ctatattagt tgtggagaag aacacacagc agttctcaca aagagtggag gtgtgtttac 960

ctttggcgct ggttcctgtg ggcaacttgg acacgactcc atgaatgatg aggttaaccc 1020

tagaagagtt ctagagctga tgggtagtga agtaactcaa attgcttgtg gcagacaaca 1080
taccctagcc ttcgtgcctt cttctggact catctatgca tttggttgtg gagcaagagg 1140
tcaattagga actgggcaca cttgtaatgt taagtgccca tctcctgtca agggttactg 1200
ggctgcccac agtggccagc tttcagcccg agctgatcgc tttaaatatc atatcgttaa 1260
gcagatcttc tctggaggag accagacttt tgtactttgc tccaaatacg agaattattc 1320
tcctgctgtt gacttcagga ctatgaacca agcacattat accagtttaa taaatgatga 1380
aaccatagca gtttggagac aaaaactctc agaacacaac aatgcaaata caatcaatgg 1440
tgttgttcag atattatctt ctgcagcctg ttggaatgga agttttcttg aaaaaaaaat 1500
tgatgaacat tttaaaacga gtcccaaaat ccctgggatt gacctgaact caactagggt 1560
gttatttgag aagttaatga actctcagca ctccatgatt ctagaacaga ttttgaacag 1620
ttttgaaagt tgtctgattc cccagttgtc aagctcacca ccagatgttg aagccatgag 1680
aatctattta atactacctg agtttcccct actccaggat tccaagtatt atataacatt 1740
gactattccc ttggctatgg ccattcttcg gctggataca aaccccagca aagtactaga 1800
taactggtgg tctcaggtat gcccgaaata tttcatgaag ctggtaaacc tctataaagg 1860
tgcagtcctt tatctactga ggggaagaaa gacattctta attcccgtac tgtttaacaa 1920
ttatatcaca gcagctctca aactcttgga gaagttatat aaggtaaatc ttaaagtgaa 1980
gcatgtggaa tatgatacat tttacattcc tgagatttcc aatctcgtgg acattcagga 2040
agactacctc atgtggttct tgcatcaagc agggatgaag gctagaccat caataataca 2100
ggatactgta acactttgtt cctacccttt catctttgat gcccaagcca agaccaaaat 2160
gttacagaca gatgctgaac tacagatgca ggtggcagtc aatggagcca acctgcagaa 2220
tgtcttcatg cttctcaccc tggagcctct gctggccaga agccccttcc tggtccttca 2280
cgttcgcagg aacaaccttg ttggagatgc cctaagagag ctgagcattc attctgatat 2340
tgatttgaaa aagcctctca aagtaatctt tgatggtgaa gaagcagtgg atgccggtgg 2400
tgttacaaag gaatttttc ttttgctgtt aaaagaactt ttgaatccca tctatggaat 2460
gtttacctac tatcaagatt caaatctctt gtggttttca gacacgtgtt ttgtagagca 2520
caactggttt cacttgattg gtataacctg tggactagct atctacaact ccactgtggt 2580
cgatctccac ttcccattgg ctctctacaa gaagttactc aatgtaaagc ctggcttgga 2640
agacttaaag gagttgtcac ccactgaagg aaggagtctc caagagcttt tagattaccc 2700
cggggaggat gtggaggaga ctttctgcct caacttcacg atctgccgag aaagctatgg 2760
agtgattgaa cagaagaagc tgatacctgg gggagataat gtaactgtgt gcaaggataa 2820
caggcaggaa tttgtggatg cttatgtgaa ttatgtcttc caaatctcag ttcatgaatg 2880
gtacacagcc ttctctagtg gcttcctaaa ggtgtgtggt ggcaaagtac ttgagctctt 2940

ccagccttca gaactgaggg ctatgatggt ggggaacagc aactacaact gggaagaact 3000

ggaagagact gccatctaca agggagatta ctcggccaca catcccactg taaaactatt 3060

ttgggaaaca tttcatgagt ttccattgga aaagaagaag aagtttctct tgttcctgac 3120

aggcagcgat cggattccca tctacggcat ggccagtctg cagattgtca tccagtccac 3180

agccagcggg gaggagtact tgccggtggc ccacacttgc tacaaccttc ttgacctccc 3240

caagtacagc agcaaagaga ttctgagtgc ccggctgacc caggcccttg acaactatga 3300

agggtttagt ttggcctgag gcttctcagc ttgtccagta tttcccttcg ttcctcagtg 3360

tccacattga ggcctataca gaaaatcatg gggagtgatt tctatttttt tattgtctaa 3420

gtgggttggg acttttaaat actgagcctg gttgatgtgt ttctgggatt gtatagcagt 3480

aaacaacctt tttgaaaaat tagaggttgg ggatggggtg aaaaattggc ccttgtatgg 3540

gaggtgtttt tgtttttgtt ttaaaccaaa ctacccagta ttccttgcac ttgtgaatgt 3600

gttgcactct gctggatgaa atggcagtgg atttttaaac tttaatttcc caaatgtctc 3660

tctcagccct gatgttttct cacagtgctt ccttgtcctt ctcttaactt ctcattcctc 3720

tataagaatg atttagactg acctgtcctt ttttatctgc gcatgcgaga acatcacctt 3780

cctctgtaca cttggaaatg cctctggctt gttgcagccc tcctttaacc caaaggagga 3840

aaggactgct tcagaaactc ccaattccaa aaagctgagt ctgggtccat tattttggca 3900

gaactcctaa gaatttatgg gagcctatat aaacatatct tgcttttaaa aagttcttga 3960

gggaatagca actttcccat ggctgtgcct atttcctaga ccttttaaaa gatgtgcaga 4020

gcagcttagc attcgttgca gctgagccta atttttttctt gctcatcctt gtccctttga 4080

caataaggtt aattgataga cccaccacct cttgcactct cgcttttgga gcaagttgca 4140

ttaactattt tgagtctcta tattgtccaa gaaaagtaga aataataaat ttactttccc 4200

tttttctatc accttatgtc ctctaccatt ttctccttcc tcccttccct tattttctcc 4260

ttttcgtacc ctgtgtcctc cctgattttc ctttcgtttc ttctttattt tatcccattc 4320

tctgttactt gactcagtgc tcccttcctc tcctctcctt ctagtggatg catgcagcct 4380

ttttttcaat ttttatttaa attgcaaaat ttttactcag atttttttttc ctcttcccta 4440

attgctaaga tttaaggacg ttctttatta tgaaacttta tcacattcga aatgtttgtt 4500

tacagtggga ttttaggggg gattgtgttt aaatcaaata tatgtatttt aaaaataatg 4560

acatgctcaa ccttcctcat catggagtaa gaaaattcta catgattaaa gaatccatgt 4620

aagtctaatt ttaaattcct agtaactaga gaaaagactt atttatataa aatgaagtat 4680

ttatgaactg tgataaagca tcaaatcttg atgaaggatt gtagattttt gctttttctt 4740

tttgttttta aaacttattc caattgctaa attggtagtt tttcagtctt tataaataca 4800

ggattaaaaa tatatataca gttatatgaa atgtttattt tctatgtgtg tgcatatagt 4860

tcaatattat gcaataaatt tggtgtttta actt 4894

<210> 28
<211> 1050
<212> PRT
<213> Homo sapiens

<400> 28

Met Leu Cys Trp Gly Tyr Trp Ser Leu Gly Gln Pro Gly Ile Ser Thr
1 5 10 15

Asn Leu Gln Gly Ile Val Ala Glu Pro Gln Val Cys Gly Phe Ile Ser
20 25 30

Asp Arg Ser Val Lys Glu Val Ala Cys Gly Gly Asn His Ser Val Phe
35 40 45

Leu Leu Glu Asp Gly Glu Val Tyr Thr Cys Gly Leu Asn Thr Lys Gly
50 55 60

Gln Leu Gly His Glu Arg Glu Gly Asn Lys Pro Glu Gln Ile Gly Ala
65 70 75 80

Leu Ala Asp Gln His Ile Ile His Val Ala Cys Gly Glu Ser His Ser
85 90 95

Leu Ala Leu Ser Asp Arg Gly Gln Leu Phe Ser Trp Gly Ala Gly Ser
100 105 110

Asp Gly Gln Leu Gly Leu Met Thr Thr Glu Asp Ser Val Ala Val Pro
115 120 125

Arg Leu Ile Gln Lys Leu Asn Gln Gln Thr Ile Leu Gln Val Ser Cys
130 135 140

Gly Asn Trp His Cys Leu Ala Leu Ala Ala Asp Gly Gln Phe Phe Thr
145 150 155 160

Trp Gly Lys Asn Ser His Gly Gln Leu Gly Leu Gly Lys Glu Phe Pro
165 170 175

Ser Gln Ala Ser Pro Gln Arg Val Arg Ser Leu Glu Gly Ile Pro Leu
180 185 190

Ala Gln Val Ala Ala Gly Gly Ala His Ser Phe Ala Leu Ser Leu Ser
195 200 205

Gly Ala Val Phe Gly Trp Gly Met Asn Asn Ala Gly Gln Leu Gly Leu

210 215 220

Ser Asp Glu Lys Asp Arg Glu Ser Pro Cys His Val Lys Leu Leu Arg
225 230 235 240

Thr Gln Lys Val Val Tyr Ile Ser Cys Gly Glu Glu His Thr Ala Val
245 250 255

Leu Thr Lys Ser Gly Gly Val Phe Thr Phe Gly Ala Gly Ser Cys Gly
260 265 270

Gln Leu Gly His Asp Ser Met Asn Asp Glu Val Asn Pro Arg Arg Val
275 280 285

Leu Glu Leu Met Gly Ser Glu Val Thr Gln Ile Ala Cys Gly Arg Gln
290 295 300

His Thr Leu Ala Phe Val Pro Ser Ser Gly Leu Ile Tyr Ala Phe Gly
305 310 315 320

Cys Gly Ala Arg Gly Gln Leu Gly Thr Gly His Thr Cys Asn Val Lys
325 330 335

Cys Pro Ser Pro Val Lys Gly Tyr Trp Ala Ala His Ser Gly Gln Leu
340 345 350

Ser Ala Arg Ala Asp Arg Phe Lys Tyr His Ile Val Lys Gln Ile Phe
355 360 365

Ser Gly Gly Asp Gln Thr Phe Val Leu Cys Ser Lys Tyr Glu Asn Tyr
370 375 380

Ser Pro Ala Val Asp Phe Arg Thr Met Asn Gln Ala His Tyr Thr Ser
385 390 395 400

Leu Ile Asn Asp Glu Thr Ile Ala Val Trp Arg Gln Lys Leu Ser Glu
405 410 415

His Asn Asn Ala Asn Thr Ile Asn Gly Val Val Gln Ile Leu Ser Ser
420 425 430

Ala Ala Cys Trp Asn Gly Ser Phe Leu Glu Lys Lys Ile Asp Glu His
435 440 445

Phe Lys Thr Ser Pro Lys Ile Pro Gly Ile Asp Leu Asn Ser Thr Arg
450 455 460

Val Leu Phe Glu Lys Leu Met Asn Ser Gln His Ser Met Ile Leu Glu
465 470 475 480

Gln Ile Leu Asn Ser Phe Glu Ser Cys Leu Ile Pro Gln Leu Ser Ser
485 490 495

Ser Pro Pro Asp Val Glu Ala Met Arg Ile Tyr Leu Ile Leu Pro Glu
500 505 510

Phe Pro Leu Leu Gln Asp Ser Lys Tyr Tyr Ile Thr Leu Thr Ile Pro
515 520 525

Leu Ala Met Ala Ile Leu Arg Leu Asp Thr Asn Pro Ser Lys Val Leu
530 535 540

Asp Asn Trp Trp Ser Gln Val Cys Pro Lys Tyr Phe Met Lys Leu Val
545 550 555 560

Asn Leu Tyr Lys Gly Ala Val Leu Tyr Leu Leu Arg Gly Arg Lys Thr
565 570 575

Phe Leu Ile Pro Val Leu Phe Asn Asn Tyr Ile Thr Ala Ala Leu Lys
580 585 590

Leu Leu Glu Lys Leu Tyr Lys Val Asn Leu Lys Val Lys His Val Glu
595 600 605

Tyr Asp Thr Phe Tyr Ile Pro Glu Ile Ser Asn Leu Val Asp Ile Gln
610 615 620

Glu Asp Tyr Leu Met Trp Phe Leu His Gln Ala Gly Met Lys Ala Arg
625 630 635 640

Pro Ser Ile Ile Gln Asp Thr Val Thr Leu Cys Ser Tyr Pro Phe Ile
645 650 655

Phe Asp Ala Gln Ala Lys Thr Lys Met Leu Gln Thr Asp Ala Glu Leu
660 665 670

Gln Met Gln Val Ala Val Asn Gly Ala Asn Leu Gln Asn Val Phe Met
675 680 685

Leu Leu Thr Leu Glu Pro Leu Leu Ala Arg Ser Pro Phe Leu Val Leu
690 695 700

His Val Arg Arg Asn Asn Leu Val Gly Asp Ala Leu Arg Glu Leu Ser
705 710 715 720

Ile His Ser Asp Ile Asp Leu Lys Lys Pro Leu Lys Val Ile Phe Asp
725 730 735

Gly Glu Glu Ala Val Asp Ala Gly Gly Val Thr Lys Glu Phe Phe Leu
740 745 750

Leu Leu Leu Lys Glu Leu Leu Asn Pro Ile Tyr Gly Met Phe Thr Tyr
755 760 765

Tyr Gln Asp Ser Asn Leu Leu Trp Phe Ser Asp Thr Cys Phe Val Glu
770 775 780

His Asn Trp Phe His Leu Ile Gly Ile Thr Cys Gly Leu Ala Ile Tyr
785 790 795 800

Asn Ser Thr Val Val Asp Leu His Phe Pro Leu Ala Leu Tyr Lys Lys
805 810 815

Leu Leu Asn Val Lys Pro Gly Leu Glu Asp Leu Lys Glu Leu Ser Pro
820 825 830

Thr Glu Gly Arg Ser Leu Gln Glu Leu Leu Asp Tyr Pro Gly Glu Asp
835 840 845

Val Glu Glu Thr Phe Cys Leu Asn Phe Thr Ile Cys Arg Glu Ser Tyr
850 855 860

Gly Val Ile Glu Gln Lys Lys Leu Ile Pro Gly Gly Asp Asn Val Thr
865 870 875 880

Val Cys Lys Asp Asn Arg Gln Glu Phe Val Asp Ala Tyr Val Asn Tyr
885 890 895

Val Phe Gln Ile Ser Val His Glu Trp Tyr Thr Ala Phe Ser Ser Gly
900 905 910

Phe Leu Lys Val Cys Gly Gly Lys Val Leu Glu Leu Phe Gln Pro Ser
915 920 925

Glu Leu Arg Ala Met Met Val Gly Asn Ser Asn Tyr Asn Trp Glu Glu
930 935 940

Leu Glu Glu Thr Ala Ile Tyr Lys Gly Asp Tyr Ser Ala Thr His Pro
945 950 955 960

Thr Val Lys Leu Phe Trp Glu Thr Phe His Glu Phe Pro Leu Glu Lys
965 970 975

Lys Lys Lys Phe Leu Leu Phe Leu Thr Gly Ser Asp Arg Ile Pro Ile
980 985 990

Tyr Gly Met Ala Ser Leu Gln Ile Val Ile Gln Ser Thr Ala Ser Gly
995 1000 1005

Glu Glu Tyr Leu Pro Val Ala His Thr Cys Tyr Asn Leu Leu Asp
1010 1015 1020

Leu Pro Lys Tyr Ser Ser Lys Glu Ile Leu Ser Ala Arg Leu Thr
1025 1030 1035

Gln Ala Leu Asp Asn Tyr Glu Gly Phe Ser Leu Ala
1040 1045 1050

<210> 29
<211> 4363
<212> DNA
<213> Homo sapiens

<400> 29
gagctggaat tcccttcccg cctggcgccg cccgggagag cccgcgcgag gacgcccgcc 60

gcgctccgcc ggccctttt tggcgctgag ggaaaggaga agggcagggc cgccgcctcg 120

cgggagcggg ccccgggagg aggggccgcc ggagaggccg ggcgagcgcg ggcggccctc 180

acctcgccgc tcctcccggg ccgccatccc tcggcgcccc gcccggaacc ggcgcgcgcg 240

taggggctgg gaggccggcg cgcagatctg gcggtgagcg ctgccgcccc ggggccccca 300

gccatgtcgg ccgaggagat ggtgcagatc cgcctggagg accgctgcta cccggtgagc 360

aagaggaagc tcatcgagca gagcgactac ttccgcgccc tctaccgctc cggcatgcgc 420

gaggccctga gccaggaggc cggcggcccg gaggtgcagc agctgcgcgg cctcagcgcg 480

ccgggcctgc ggctggtgct ggacttcatc aacgccggcg gggcccgcga aggctggctc 540

ctgggcccgc gcggggaaaa gggcggcggg gtggacgagg acgaggagat ggatgaggtg 600

agcctgctgt ccgagctggt ggaggcggcc tccttcctgc aggtcacgtc cctgctgcag 660

ctgctgctgt cccaggtgcg gctcaataac tgcctggaga tgtaccgcct ggcgcaggtg 720

tacgggctgc ccgacctgca ggaggcctgc ctgcgcttca tggtcgtcca cttccacgag 780

gtgctgtgca agccccagtt ccacctcctg gggtctcctc cccaagctcc aggggatgtc 840

agcctgaagc agaggctgag ggaggcccgg atgactggga ctcctgtcct cgtggccctc 900

ggggacttcc tgggggggacc cctggcccct caccctacc agggggagcc cccgtccatg 960

ctcaggtacg aggagatgac tgagcgttgg ttcccgctgg ccaacaacct tcctcccgac 1020

ctggtcaatg tcaggggcta tgggtctgcc atcctggaca actacctctt catagtgggc 1080

gggtacagga tcactagcca ggagatctcc gctgcgcatt cctacaaccc cagcaccaac 1140

gagtggctcc aggtggcctc catgaaccag aagaggtcta acttcaaact tgtggctgtt 1200

aattcaaaac tctatgccat cggagggcag gccgtttcta acgttgagtg ttacaacccc 1260

gagcaggatg cgtggaattt tgtggcgccc ttacccaatc ctctggctga gttctctgcc 1320

tgtgagtgta agggaaaaat ttatgtcatt ggaggataca ctaccagaga ccggaacatg 1380

aacattttgc agtactgccc ctcttccgac atgtggacgc tctttgaaac atgtgacgtc 1440

cacattcgca agcagcagat ggtgtctgtg gaagagacca tctacatcgt gggggggtgt 1500

ctccacgagc tggggcccaa ccgcaggagc agccagagcg aggacatgct caccgtgcag 1560

tcctacaaca ccgtcacccg ccagtggctc tacctcaagg agaacacgtc caaatcgggt 1620

cttaacttga cttgtgcgct ccataacgac ggcatctaca tcatgagcag agacgtcacc 1680

ctgtcgacca gcttggaaca ccgagtgttc ctcaagtaca acatcttttc agatagttgg 1740

gaagcatttc ggcgttttcc agcttttgga cataacttgc tggtttcttc tctttatctg 1800

cccaataaag cagaaacatg actgaattga attggtagat gaaaaaaacc tggttcaagt 1860

ttttttaaa atgtggtgtc ccattccaag ggagaccaat tcctaaaggg taaagaaggg 1920

ttaaagtagg tcacatatat acagtagcag ctgtaaataa gcttacttga actaattact 1980

tgaaaactgg tggaaaaaag agaccaactt tgttattttt taaattattg atttttaaat 2040

tatgggagca aatccatctc caaacatgat actcttgggc caatgacgga tcatcaaaat 2100

gtagattcat tgtgtctgtg tggtaaaggg ccaaagtcaa taattgggac aaatggtaaa 2160

gatgatttta aaatctgttg tatttaggtc ccttgacata tcattagcac cttaagtgag 2220

aggttttctt tttcttatca gtgtccttga caagcattca tattaacata ttcattctct 2280

cttctccatc acccataccc acacatacat ctagtatgtt tagggttctg tgttcagata 2340

ttttgtttgc acactacttt ttaggaaact tataatgatg gagccctaca gtgaatcctt 2400

accagtgttc agtcctcccc aaaacttgca taaaaatatt cttagctatt tttccagcct 2460

tttttgcagt tacaaatgct ggataaatag actgataaaa cctcttattt tatttatgtg 2520

gcaggttgca tatttatgtg gacacatagg ttgcactgct aggcattctg tattgtttta 2580

aagaggatat attatttaga aaaatggaca tatgaatatt attaaaaatt gagtcttaaa 2640

tttaaatggg aagggaagaa aacaccaaat aagaaaaaaa tttgccttta agagttactt 2700

ttgttgaatg tatttgactt agttcttatt ttaaggctaa aaaaactttg tcatatttaa 2760

gtcggcagct ttctaattta atcttagctt tgtaattcat gttttgcttt tgtggactaa 2820

ggtaaggaca gcatttaaat atttgtgctg gagtgagggg aagaagctgt tacagaagtg 2880

gaatggtttc tggtggtatt tctgaagtta ggtaaagtgc tttacttgcc ttggttgtta 2940

attgaaatag aattcagctc aagcagttgt tttcagtgag tgagttaaca ataaaagctt 3000

ctgccaatca tagatcaaca ttcctcaaaa atacatacac tcatacttat ggaaatttta 3060

ttttgattga aaacactctg aaattatctg atattttgta atggcatcag ttaaactgtg 3120

agctggattc catctttctg cagtaacagc gggcagaacc atttcagtta gcatgtgcct 3180

cactttaaga aagtcctgta ggggaaaaaa aaaactgtca gagcgaattc atgcattcaa 3240

caaacactta tgagtgcctg gagtatgccg tccactgtgc agatgtagga caggggctaa 3300

tcacttagag ttcttcagtt tattagagga gtgactacag gttttcttta caaccatggc 3360

ctctagtata cttaaaataa gcgtccagtt ataaaaagcc tatttgcaca acttttgggg 3420

aatacttcta ttaactaaaa tgaggtaagc cattattttg gacatagaaa cagatctttg 3480

tatcttgtca ccaaataatc ttacctctag tctacttaca actttgtaag ggtttctaag 3540

ctatgcccat ttattcccaa atattaagcc cttaaaaaaa taaaacctct gaaccaaaat 3600

cttcccagga atagtactta atagatttta atgttaattt atattcattg tagaattgaa 3660

tttccagtga cctgtaattt cattacatat attttaaaat ctatttattt tagatggtgt 3720

taacttttcg ttgctgaaga gcttaacttt ttaaaggctt tgtcctatac atttagaaga 3780

tgagtcagtt gaattagtac tggacaaaca gttggagatt agtttgcaaa taagcatatc 3840

atatcaaata caatattcca gccaaccagt taattctctt cctggttaca ttcattgggt 3900

ctatttgcct aatgttgaca attagattct tggaccaagt gagggtgtac tctgatccac 3960

aaagctcttt tataaaccag gtttgaggtt gtgtgtgtgt gtgggggtgt gtgtgtgtgt 4020

gtgtgtgtat gtttgcatat tatagctctc tttaagacag acatctaaaa agattttggt 4080

tattctgggc tgccagcagt acgtgggaaa ggtggaatgt gcaataggaa cagatctgtg 4140

cattttgtga catcgtcctt cctgctgtgc acatcagctg tgatgacatt ttgagggtag 4200

ttttaattta gctcggttct cttgttggtg ccagtattac attgaagact acctatgata 4260

gaggcaatgt tgctttgaaa atctcccttc aggcggaggt tgcagtgagc cgagattgtg 4320

ccattgcact ccagcctggg cgacagtgca agactctgtc tcc 4363

<210> 30
<211> 505
<212> PRT
<213> Homo sapiens

<400> 30

Met Ser Ala Glu Glu Met Val Gln Ile Arg Leu Glu Asp Arg Cys Tyr
1 5 10 15

Pro Val Ser Lys Arg Lys Leu Ile Glu Gln Ser Asp Tyr Phe Arg Ala
20 25 30

Leu Tyr Arg Ser Gly Met Arg Glu Ala Leu Ser Gln Glu Ala Gly Gly
35 40 45

```
Pro Glu Val Gln Gln Leu Arg Gly Leu Ser Ala Pro Gly Leu Arg Leu
     50                  55                  60

Val Leu Asp Phe Ile Asn Ala Gly Gly Ala Arg Glu Gly Trp Leu Leu
65                  70                  75                  80

Gly Pro Arg Gly Glu Lys Gly Gly Gly Val Asp Glu Asp Glu Glu Met
                85                  90                  95

Asp Glu Val Ser Leu Leu Ser Glu Leu Val Glu Ala Ala Ser Phe Leu
            100                 105                 110

Gln Val Thr Ser Leu Leu Gln Leu Leu Leu Ser Gln Val Arg Leu Asn
        115                 120                 125

Asn Cys Leu Glu Met Tyr Arg Leu Ala Gln Val Tyr Gly Leu Pro Asp
    130                 135                 140

Leu Gln Glu Ala Cys Leu Arg Phe Met Val Val His Phe His Glu Val
145                 150                 155                 160

Leu Cys Lys Pro Gln Phe His Leu Leu Gly Ser Pro Pro Gln Ala Pro
                165                 170                 175

Gly Asp Val Ser Leu Lys Gln Arg Leu Arg Glu Ala Arg Met Thr Gly
            180                 185                 190

Thr Pro Val Leu Val Ala Leu Gly Asp Phe Leu Gly Gly Pro Leu Ala
        195                 200                 205

Pro His Pro Tyr Gln Gly Glu Pro Pro Ser Met Leu Arg Tyr Glu Glu
    210                 215                 220

Met Thr Glu Arg Trp Phe Pro Leu Ala Asn Asn Leu Pro Pro Asp Leu
225                 230                 235                 240

Val Asn Val Arg Gly Tyr Gly Ser Ala Ile Leu Asp Asn Tyr Leu Phe
                245                 250                 255

Ile Val Gly Gly Tyr Arg Ile Thr Ser Gln Glu Ile Ser Ala Ala His
            260                 265                 270

Ser Tyr Asn Pro Ser Thr Asn Glu Trp Leu Gln Val Ala Ser Met Asn
        275                 280                 285

Gln Lys Arg Ser Asn Phe Lys Leu Val Ala Val Asn Ser Lys Leu Tyr
    290                 295                 300
```

```
Ala Ile Gly Gly Gln Ala Val Ser Asn Val Glu Cys Tyr Asn Pro Glu
305                 310                 315                 320

Gln Asp Ala Trp Asn Phe Val Ala Pro Leu Pro Asn Pro Leu Ala Glu
                325                 330                 335

Phe Ser Ala Cys Glu Cys Lys Gly Lys Ile Tyr Val Ile Gly Gly Tyr
            340                 345                 350

Thr Thr Arg Asp Arg Asn Met Asn Ile Leu Gln Tyr Cys Pro Ser Ser
        355                 360                 365

Asp Met Trp Thr Leu Phe Glu Thr Cys Asp Val His Ile Arg Lys Gln
    370                 375                 380

Gln Met Val Ser Val Glu Glu Thr Ile Tyr Ile Val Gly Gly Cys Leu
385                 390                 395                 400

His Glu Leu Gly Pro Asn Arg Arg Ser Ser Gln Ser Glu Asp Met Leu
                405                 410                 415

Thr Val Gln Ser Tyr Asn Thr Val Thr Arg Gln Trp Leu Tyr Leu Lys
            420                 425                 430

Glu Asn Thr Ser Lys Ser Gly Leu Asn Leu Thr Cys Ala Leu His Asn
        435                 440                 445

Asp Gly Ile Tyr Ile Met Ser Arg Asp Val Thr Leu Ser Thr Ser Leu
    450                 455                 460

Glu His Arg Val Phe Leu Lys Tyr Asn Ile Phe Ser Asp Ser Trp Glu
465                 470                 475                 480

Ala Phe Arg Arg Phe Pro Ala Phe Gly His Asn Leu Leu Val Ser Ser
                485                 490                 495

Leu Tyr Leu Pro Asn Lys Ala Glu Thr
            500                 505
```

<210> 31
<211> 2501
<212> DNA
<213> Homo sapiens

<400> 31
gttgctggct ttttcaggag ctgctcgctc gcagccagag acgctgcttt ttttttccgg 60

gttcggagcc gttccggatg ctttaggctg ccggatgtct gatctccgaa taactgaggc 120

gtttctgtac atggattatc tgtgttttag agcactttgc tgcaagggac caccacctgc 180

acgaccagaa tatgacctgg tttgcatagg cctcacaggt tctggcaaaa ccagtctgtt 240

gtccaaactc tgcagtgaaa gccccgataa cgtcgtgtcg accacaggtt ttagtattaa 300

agcagtgcca ttccagaatg ccatcttgaa tgtaaaagaa cttggagggg ctgataacat 360

ccggaaatac tggagccgct actaccaagg atctcaaggg gtaatatttg tattagacag 420

tgcctcttca gaggatgatt tagaagctgc tagaaatgag ctgcactcag ctcttcagca 480

tccacagtta tgcactttac ccttttttaat attggccaat catcaagaca agccagcagc 540

tcgctcagta caagagatca aaaaatattt tgaacttgaa ccacttgcac gtggaaaacg 600

ctggattcta cagccctgct cactggatga catggatgca ctgaaagaca gcttctctca 660

gctgattaat ttgttagaag aaaaagacca tgaagctgta agaatgtgaa atctggcaaa 720

gaaaacaggc tcccaaaagg ccttgaacct atcatgttag tttctcattt taattttatt 780

ttggtatcaa gactatattg gcttcagtat ctattttcct ctggtcattt cagactctca 840

tctctctatt aaagtgaata ttctgtagat cagatgaatt atcatcggca ttaacgtcaa 900

gtacacacta ctgagagaga atttattctc tgtttaccac taattatctt ccccgcatgt 960

ctcttccttc tgttgttcta agcatctctt ctatactctg tgctgaaacg gatggaattc 1020

atggcttatg gaaatcaaat ccctttttgaa ggactcagaa agactcacaa gaccttgatc 1080

attttctctg ggtttgggag tggagtatta tgtgactggg actccacaaa atatttatat 1140

tattttcaag acctattggg ggttttgttt tattttatct aatttttttt ttttttggt 1200

atactgagct tatagactat gagagaaaag tgagcaccaa aaacttcatt aatcacctta 1260

tcaggaagga atatcctgcc atagtaaaat tggctctgaa gagcttagag gcaagttaca 1320

caatgttttc ttgattacat ctgaagtctc ttctctttta agagcaaaga cattgtttca 1380

gtttctgaga gtggttagag gaaatttagg tacaatgcat gcatcacaga ctgtatttga 1440

ttgattcgag tcttcaaata ttgtttaatg gaatgatgag cctgcataaa gtgaagctgc 1500

aatgtaacat aggtatcttt ttctaaaaac ttcaagtgtt tgattaaaat agagaaaata 1560

atttttaaaa taatagataa aaattttaaa aacagccaga aaaagtataa tttatactgg 1620

ttagggttta agatggcatg gatttgacaa agataacaaa gacagaatca tctatatatt 1680

cattcattac gtgttactgt agtttattgg aatctgttaa tgccttcctc tagtagttaa 1740

aggtttatca gtatttctat tataaacctc tattttcagg ggtttcaatg tgcccttaaa 1800

attgtgctgg ggctgcagca aataaataat attgggactc atattcagtc tttggaacta 1860

gtttttttttc catttaattt aaatgacaaa attgcatagg gataaatgta cagttttact 1920

aaggtttagc acctttgatc ttttctaaat caggaaaaaa atctatgcat attgaggttc 1980

acagatgaaa tattctttgg catactttgt agactttgac cacttaaaat actaagtggc 2040

caaatgataa aaaaaactgc tccctgcaca tgagtgtaaa aacatgtttg taattttaag 2100

taaaatggac tgatacccaa ataccacaga ctcacactat attacagaat gatgtaatac 2160

ctggtaattc tccagaagat agacttcaga cccttaatgc catttaaata tagtttttgt 2220

ggtttaatct tatgagtaga ttgtaatgaa atcattacct ttgaaacatc aaacacttct 2280

ttaagctttt ttttttaat gtcatgaatc aggaaggata catagtgatt cgtatacaat 2340

tttgtctgca gcattcttga atatttttaa cactgtgtaa tccctgagca gcacaaaact 2400

tttattccga tgaccagtta cacaattcct gtttttattt tgtcctgttg tgctcatctt 2460

ctatgaaaaa taaatgatta ttaggcaaaa aaaaaaaaaa a 2501

<210> 32
<211> 204
<212> PRT
<213> Homo sapiens

<400> 32

Met Ser Asp Leu Arg Ile Thr Glu Ala Phe Leu Tyr Met Asp Tyr Leu
1 5 10 15

Cys Phe Arg Ala Leu Cys Cys Lys Gly Pro Pro Pro Ala Arg Pro Glu
20 25 30

Tyr Asp Leu Val Cys Ile Gly Leu Thr Gly Ser Gly Lys Thr Ser Leu
35 40 45

Leu Ser Lys Leu Cys Ser Glu Ser Pro Asp Asn Val Val Ser Thr Thr
50 55 60

Gly Phe Ser Ile Lys Ala Val Pro Phe Gln Asn Ala Ile Leu Asn Val
65 70 75 80

Lys Glu Leu Gly Gly Ala Asp Asn Ile Arg Lys Tyr Trp Ser Arg Tyr
85 90 95

Tyr Gln Gly Ser Gln Gly Val Ile Phe Val Leu Asp Ser Ala Ser Ser
100 105 110

Glu Asp Asp Leu Glu Ala Ala Arg Asn Glu Leu His Ser Ala Leu Gln
115 120 125

His Pro Gln Leu Cys Thr Leu Pro Phe Leu Ile Leu Ala Asn His Gln
130 135 140

Asp Lys Pro Ala Ala Arg Ser Val Gln Glu Ile Lys Lys Tyr Phe Glu
145 150 155 160

Leu Glu Pro Leu Ala Arg Gly Lys Arg Trp Ile Leu Gln Pro Cys Ser

165 170 175

Leu Asp Asp Met Asp Ala Leu Lys Asp Ser Phe Ser Gln Leu Ile Asn
180 185 190

Leu Leu Glu Glu Lys Asp His Glu Ala Val Arg Met
195 200

<210> 33
<211> 2334
<212> DNA
<213> Homo sapiens

<400> 33
ctttatggcg tgggagaggc cacagcccgg actccatcga ctccccggc tcttagacta 60

aaatcatgcc caagttcaaa caacgaagac gaaagctaaa agccaaagcc gaaagattat 120

tcaaaaaaaa agaagcctct cactttcagt ccaagctaat tacacctcct cctccaccac 180

cctcaccaga aagagtcggt atttcttcaa tagatatatc tcaaagcaga agctggctaa 240

catcatcctg gaacttcaat tttcctaaca tcagagatgc aataaaactt tggacaaata 300

gagtatggtc tatatacagc tggtgccaga actgcataac ccagagttta gaagtattga 360

aagacaccat ctttccatct cgtatctgcc accgagaact ttacagtgta aaacaacagt 420

tttgcatttt ggaaagtaaa ttatgcaagc ttcaggaagc actgaagacc atctcagaaa 480

gttcttcctg tccaagctgt ggtcaaacat gtcacatgag tggtaaactt acaaatgtgc 540

ctgcctgcgt tctgatcacc cctggagact ccaaagctgt gcttcctccc acactgccac 600

agccagccag ccattttcct cctcctcctc cacctccacc tctgccacct cctccaccac 660

cactagcacc tgtgttgctc agaaaaccca gtctcgctaa agcacttcag gctggaccat 720

taaaaaaaga tggacccatg cagataacag ttaaagatct gctgactgta aaattaaaga 780

agacacagag tttagatgaa aagaggaagc ttataccatc gccgaaagca cggaatccac 840

tagttaccgt ctctgacttg cagcatgtta ccctgaaacc taactccaaa gtgttatcga 900

ctcgagttac aaacgtctta atcactcctg gaaaaagtca gatggatctg cggaaactgc 960

ttagaaaagt cgatgtagag aggaggccag gtggaacccc tcttaccaat aaggaaaata 1020

tggaaacagg aacaggactg actccagtga tgacgcaggc cttaaggaga aagtttcagc 1080

tggctcaccc tagaagccca actccaactc tgccactttc tacaagcagc tttgatgaac 1140

aaaactgatg ccaactctgc ctcactccat gagatgatcc ataacaaata cagataaaaa 1200

cacccagctg ggtgcagtgg ctcgcacctg taatcccagc actttgggag gctgaggcag 1260

gtggatcacc tgaggtcggg agtttgagac cagcctggcc aacatggtga aatcccatct 1320

ctactgaaaa tacagaattg gccgggtgtg gtgtcacgcg cctgtggtcc cagttacttg 1380

aacctgggag gcggaggttg cagtgagccg agattgcaac actgccactc cagcctgggt 1440

gacagagcaa gagaccctgt ctcaaaaaaa aaaaaaaaaa aattaaatgg gtagtgacgt 1500

taagagatat atatcagctt ctagtaaaag ttttttttt taaacctgct agctacattt 1560

acattatgta aaaataaatg gaataatcac tgagaataaa gcagttgagt atttataaca 1620

ataatatttt atggggcgct tataatgttt ataatattgt aaaccactgt gtactctatt 1680

catttaatgc taaatgactt gaccattctt gtgggataag agatcattaa aaaaatgcta 1740

gggccgggca ccatggctca cgcctgtaat cccaggactt taggaggcca aggcaggtgg 1800

atgacttgag ctcaggagtt tgagaccagc ctgggcaaca tgatgaaaac tccgtctcta 1860

ccaaaaatca aaaaaattag ccaggtgtga tgttgtgcgc ctgtagtccc agctacttgt 1920

gaggctgagg tgggaggatg gtttgggccc cggaggtaga ggttgcagtg agctgagatt 1980

gtgccactgt attccatcct gggcaacaga gccagaccct gcctcaaaaa aaaaaaaaaa 2040

aaaaaaaagg cctggcacag tggctcacgc ctgtaatccc agcactttgg gaggctgagg 2100

cgggcaaaac acttgaggtc agacgtttga gaccagcctg gccaacatgg tgaaaccccg 2160

tctctaccaa aaataaaaat aaaaattatc caggcgtggt ggcatgtgcc tatagcccca 2220

gctactcagg aagctagagc acgagaatta tgtgcatcct ggaggtgaag gttgcagtga 2280

gccaagatca caccactcca ctccagtctg gacaacagag tgagaccctg tgtc 2334

<210> 34
<211> 360
<212> PRT
<213> Homo sapiens

<400> 34

Met Pro Lys Phe Lys Gln Arg Arg Arg Lys Leu Lys Ala Lys Ala Glu
1 5 10 15

Arg Leu Phe Lys Lys Lys Glu Ala Ser His Phe Gln Ser Lys Leu Ile
20 25 30

Thr Pro Pro Pro Pro Pro Pro Ser Pro Glu Arg Val Gly Ile Ser Ser
35 40 45

Ile Asp Ile Ser Gln Ser Arg Ser Trp Leu Thr Ser Ser Trp Asn Phe
50 55 60

Asn Phe Pro Asn Ile Arg Asp Ala Ile Lys Leu Trp Thr Asn Arg Val
65 70 75 80

Trp Ser Ile Tyr Ser Trp Cys Gln Asn Cys Ile Thr Gln Ser Leu Glu
85 90 95

Val Leu Lys Asp Thr Ile Phe Pro Ser Arg Ile Cys His Arg Glu Leu
100 105 110

Tyr Ser Val Lys Gln Gln Phe Cys Ile Leu Glu Ser Lys Leu Cys Lys
115 120 125

Leu Gln Glu Ala Leu Lys Thr Ile Ser Glu Ser Ser Ser Cys Pro Ser
130 135 140

Cys Gly Gln Thr Cys His Met Ser Gly Lys Leu Thr Asn Val Pro Ala
145 150 155 160

Cys Val Leu Ile Thr Pro Gly Asp Ser Lys Ala Val Leu Pro Pro Thr
165 170 175

Leu Pro Gln Pro Ala Ser His Phe Pro Pro Pro Pro Pro Pro Pro Pro
180 185 190

Leu Pro Pro Pro Pro Pro Pro Leu Ala Pro Val Leu Leu Arg Lys Pro
195 200 205

Ser Leu Ala Lys Ala Leu Gln Ala Gly Pro Leu Lys Lys Asp Gly Pro
210 215 220

Met Gln Ile Thr Val Lys Asp Leu Leu Thr Val Lys Leu Lys Lys Thr
225 230 235 240

Gln Ser Leu Asp Glu Lys Arg Lys Leu Ile Pro Ser Pro Lys Ala Arg
245 250 255

Asn Pro Leu Val Thr Val Ser Asp Leu Gln His Val Thr Leu Lys Pro
260 265 270

Asn Ser Lys Val Leu Ser Thr Arg Val Thr Asn Val Leu Ile Thr Pro
275 280 285

Gly Lys Ser Gln Met Asp Leu Arg Lys Leu Leu Arg Lys Val Asp Val
290 295 300

Glu Arg Arg Pro Gly Gly Thr Pro Leu Thr Asn Lys Glu Asn Met Glu
305 310 315 320

Thr Gly Thr Gly Leu Thr Pro Val Met Thr Gln Ala Leu Arg Arg Lys
325 330 335

Phe Gln Leu Ala His Pro Arg Ser Pro Thr Pro Thr Leu Pro Leu Ser
340 345 350

Thr Ser Ser Phe Asp Glu Gln Asn
355 360

<210> 35
<211> 6743
<212> DNA
<213> Homo sapiens

<400> 35
cgacggcggc ggccgtcctc atcccggcgc ttgagaggac gcggggctgc gcaaatggct 60
tgtcccgctc tcggtctgga agctcttcag cccctgcagc ccgagccgcc ccccgagccc 120
gccttctccg aggcgcagaa gtggattgag caagtaactg gcagaagttt tggtgataaa 180
gattttcgga caggtttaga aaatggaatc ctcctctgcg agttgctgaa tgctataaag 240
ccaggacttg ttaaaaagat caatagattg cctaccccca ttgcaggact ggacaatatt 300
atcttattct tgagaggttg taaagagctc ggccttaaag aatctcaact ttttgacccg 360
agtgacctcc aggatacatc caacagagta acagtcaaga gccttgatta tagtaggaag 420
ctgaaaaatg tattagttac catttactgg ctgggaaaag cagcaaacag ctgcacatcc 480
tacagcggaa cgacactaaa cctgaaggag tttgaaggat tgttggctca gatgcgaaag 540
gacactgatg acattgaaag tcctaaacgc agtatccgag acagtggcta catcgactgc 600
tgggattccg agcgcagcga ctccctctct cctcctcgcc acggcagaga tgattccttc 660
gacagcctgg attcctttgg ctctcgctct cggcagacgc cttcaccaga tgtagtcctc 720
aggggaagca gcgatgggag aggaagcgac tctgaatccg acttgcctca tcggaagctg 780
ccagatgtga agaaggatga catgtctgca cggcggactt cccatggtga gccgaaatca 840
gcagtgcctt ttaaccagta cctcccgaac aaaagcaatc agacggccta cgtccccgcg 900
cctctgagaa agaagaaagc agagagagag gaataccgca agagctggag taccgccacc 960
tccccgctgg gtggggagag gcccttcagc tttcctgaaa cgatagagga agaggggagt 1020
gaagtggggt ctgctggtga agacaaccca gctggccaaa tgaatcctgg ttggaagcct 1080
tccgatggtg ggtgtgaatt acctgacggc agtggtaagg agcacccttc ttcagacggg 1140
gctgtggtgg caccagctcc caagtctgaa gaaaaagatg ctgctgagat ccaaaagcgc 1200
aaaaggctag agcaagctgg aatcaaggtc atgccagcag cacagcgctt tgccagtcaa 1260
aaacaattaa gtgaagagaa agaagctatt cgtgatattg tccttcgcaa agaaaactct 1320
ttcctgaccc accaacatgg caacgattca gaggcagaag gtgaagttgt gtgtcgactg 1380
cctgatcttg agaaggatga ctttgctgca aggagagcaa ggatgaacca aaccaagcca 1440
atggtgccat taaatcaact cctctatggc ccttatccaa agaaaggggc agagaaatca 1500
gatggcagca aacagctctc aaagggaatc tcaaaaaaaa gaagtctaga atataaaaga 1560
aaccagggcc acacagaaga ggtgaagttg atagtgacct gtaacatgag ggctcaggaa 1620

agtgaacctg tggaaggagg actcaggaag gtgccagatc ttcacaagga tgacctggcc 1680
cagcagagaa ttcagggcag ccttgcccct caccgcgagc ccccgagctt cattacgctc 1740
tccaacataa cagaagctga cttggagacg tgggagagac tcaaagtgtc agaaaaggcg 1800
agggatggag atgttcagca catctgtgct tctgagcctt ccccagaaat taaagcagaa 1860
actgccattc gtgatgactt tgccaaccgc aaagcaaggg cctctaagaa agcttccagc 1920
cccaggcaaa agtttgtgca ctttgggcca gtgacggagc tagatcagca gaaatggaag 1980
cggcttagca ttggcaaggc tgggcctaga gaggatgaag aagaagtcat ctgtcatggc 2040
agcaagattc aaatggactc tgtgtctcct gtctcagcgg ccacttccag cttaaagggc 2100
caccaaatat ttaatcgaca aaatgactgc agaacaatga attgtggccg aggtgactat 2160
tgcagaaggg cctcgtggct ggctcctgtg ccggagtctc aggaagagtg ggtctgcagt 2220
ttgggcgagt gcccgagggg gacagaggaa gttacatcca aacagctgcc acaggatggc 2280
aaagaagaaa cagaaagcgc tccaagagat tctgagaggc tgtcaaaggc agaaagatca 2340
gaggacagca gccagccact ggtgtgtcct ctggcctctg agtgtgaggc ttcagggaca 2400
gaagagaagt tggagaagat gacagctcct gcctggagtg gcagtggact gaaaggccag 2460
cgaaagttgg atgattcacg aaaagatgac atgatggcca ggagaactgg gatgtccctt 2520
agacacactg gatccaaccc aaaccagttc cttccagttc catttgcaaa gcaacaggat 2580
gtggaagaat cttctaaagg tctacccatg aaagatcaga gatacggtcc gagaactcct 2640
gtgtctgatg acgcagagag caccagcatg tttgacatgc ggtgtgagga ggaggccgcg 2700
gtgcagccgc acagcagggc ccgccaggag cagctgcagc tgataaataa ccagctgagg 2760
gaagaggacg acaaatggca agatgacctg gctcgttgga agagtcgtag aagaagtgtt 2820
tctcaggact taatcaagaa agaggaagaa aggaaaaaaa tggagaagtt actggctgga 2880
gaagatggga caagtgaacg aaggaaaagc atcaaaacct acagagaaat tgttcaagaa 2940
aaagagcgga gagagagaga gctgcatgaa gcatataaga acgctcggtc ccaggaggag 3000
gcagagggga tccttcaaca gtacattgag aggttcacca tcagtgaggc tgttctcgaa 3060
cgcttggaga tgccaaaaat tctggaaaga agccattcaa cagagccaaa tttatcctcc 3120
ttcctgaatg accccaatcc catgaaatac ctgcggcaac agtcactgcc tccacccaaa 3180
ttcactgcca ctgttgaaac caccattgct cgtgccagtg ttctggatac cagcatgtca 3240
gcaggcagtg ggtctccaag caaaactgtc actcccaaag cagtgcctat gctgacaccc 3300
aagccttact cccagcccaa aaattctcaa gatgttctga agacctttaa ggtagacggg 3360
aaagtcagtg tgaatggaga gacggttcat agagaggagg agaaggaaag agagtgtccc 3420
acggtggcac ctgcccactc cttaaccaaa tcccagatgt ttgaaggtgt ggccagagtg 3480
cacgggtctc cactggagct gaaacaagac aacggtagca tcgagatcaa cataaagaag 3540

ccaaactctg ttccccaaga gctcgcagca accactgaga aaacggaacc gaatagtcaa 3600
gaggacaaga atgatggtgg aaaatcaaga aaagggaata tagaacttgc ctcatcagaa 3660
ccacagcatt ttacaacaac tgtgactcga tgcagcccga ccgtggcctt tgtggaattt 3720
ccctccagcc cccagctgaa gaatgatgtg tcggaagaaa aagaccagaa gaaaccagaa 3780
aatgaaatga gtggaaaggt ggagttggtg ctgtcacaaa aggtggtaaa gccaaaatct 3840
ccagaacccg aagcaacgct gacatttcca tttctggaca aaatgcctga agccaaccaa 3900
ctacatttgc caaatctcaa ttctcaagtg gattctccaa gcagtgagaa gtcacctgtt 3960
atgacacctt ttaagttctg ggcatgggac ccagaagagg agcgcaggcg acaggaaaaa 4020
tggcaacagg aacaggaacg tttgctccag gagagatacc agaaggagca ggacaagctg 4080
aaagaagagt gggaaaaggc ccaaaaggag gtggaagagg aagaacgcag atactatgag 4140
gaggagcgta agataattga agacactgtg gttccattta ctgtttcttc aagttccgct 4200
gaccagctgt ctacctcttc ctccatgact gaaggcagtg ggacaatgaa taagatagac 4260
ctgggaaact gtcaagatga aaaacaagac agaagatgga agaaatcatt ccagggagat 4320
gacagtgact tattgctgaa gactagggaa agtgatcgac tggaggagaa gggcagccta 4380
actgaagggg ccttggctca ttctgggaac cctgtatcaa aaggagtcca tgaagaccat 4440
cagctggata ccgaggctgg ggccccacac tgtggaacaa acccacagct tgctcaggat 4500
ccatcccaga atcagcagac atcaaatcca acgcacagtt cagaagatgt gaagccaaaa 4560
accctcccgc tggataaaag cattaaccat cagatcgagt ctcccagtga aaggcggaag 4620
aaaagtcccc gagagcactt ccaggctggg cctttctctc cctgttctcc cacccctccc 4680
ggtcagtcac caaacaggtc tataagtgga aagaagctgt gctcttcctg tgggcttcct 4740
ttgggtaaag gagctgcaat gatcatcgag accctcaatc tctattttca catccagtgt 4800
ttcaggtgtg gaatttgtaa aggccagctt ggagatgcag tgagtgggac ggatgttagg 4860
attcgaaatg gtctcctgaa ctgtaatgat tgctacatgc gatccagaag tgccgggcag 4920
cctacaacat tgtgacacgg ctttcaagct tccggatcac tcaccatttc tttactgaga 4980
gtgtcccctg gcaactgctt aacaaaatcc caagctcagg ggcttctcag catttaccta 5040
atttctgaaa ggctcttctg aaaggtggta tctgttcttt cgtagcacag tgtttatgtt 5100
tttcctgttt attgttttgg ttttttttct ttttttgcat ttgcacagta tacacaaaag 5160
aatatggggt tgtaatgatc ctgaatagct caaaaaaggt tttagcatgg tcaaacaggc 5220
ttatggttta aaatgtgtta ttctcttctt tgggaattag ctaaatgatg caataaacct 5280
gttttgtttt agaatgtcta ggaattaaac actttatgtt tacagaattg agctgcagaa 5340
agtgcaagac atgccaattt gagacacacg gtcttctaag actgaaggat aaatttaatg 5400
catttcagaa actaaacatc acagcaagct ctatctctga gctataattt gtttttaatg 5460

caaagacact agtttgataa tatatactgt aatcctgaaa catttgtgtt acttaccttt 5520

ggaggtagaa attataccaa taaattattg caccgttagt attagattct gtgtaccttg 5580

gaagttatgt cattaatata ggctggttca tcaaataaag caaaaccttg caatatcagc 5640

tagatttaca ctccgggacg ttgcccaaag gtaggaagaa agcagaggga aatatttcag 5700

tcatcatttc caaagtcatt atcaaaatct gtgaggaagt ttaatcttcc aaagagtcaa 5760

tgtcagacat caggcctctg ttgcctgctt ctctcgaggc actagattag gagtcttcaa 5820

taagagactt aacatgaggt atatggaaga tgaggcaccg agataagttc atcattaggt 5880

gtgagcactg ctcacccttg ctggcaagtt ctccttaagg gcctgaagca caggtgtcca 5940

aagaaaagcg ttaagtccat cttaatagaa tctatgtggt atatgatgtg gtcagcccct 6000

ggtctgtgat cagcaagaac ctacagcaca gattatgccc tgcccacttc aatgaatacc 6060

tactctcctc cattctccat cacttttttt gctatcaaga actccggacc ttgcccatgg 6120

agaagtttag agaggaactc ttgtggagag ctggtttatt ttctgccctg tgcgacgagt 6180

ttcagctggc caagaaagga gtcaagttat taaaaagcat cacaatgtag atctccaggc 6240

tggtttttg ttttttgttg ttaagactgg ggaaaggggg actatttatt ctgccttaaa 6300

tcaatggcaa ataagtcaag atgacatttt gtgaatgtag actatggata cactcctaat 6360

agattgatgt agtcataaaa gggggtcaag tagatgtttt tctgttatgt aagcaataat 6420

ttttccgtgt cttattgagt atggctagcg attatttatt acatgctaga tgggttcttt 6480

gcatgtgggt tccatatagg tgcagaaatt tcctcagcca ctggagggat ttcgaccata 6540

tttgtcattt ggatgagctg ttattagatt gaaatctaca catcatttca ttaaaaattg 6600

tgccttagaa aacgcaaagc tgttgcacat ggcgataaat tatggatgca gtacattgaa 6660

gagagatgaa gtcacttcca agtttccaag acttctcatg gaggtgtttg ctgttttaca 6720

ggaaaaaata aaaataaaaa aag 6743

<210> 36
<211> 1467
<212> PRT
<213> Homo sapiens

<400> 36

Met Arg Lys Asp Thr Asp Asp Ile Glu Ser Pro Lys Arg Ser Ile Arg
1 5 10 15

Asp Ser Gly Tyr Ile Asp Cys Trp Asp Ser Glu Arg Ser Asp Ser Leu
20 25 30

Ser Pro Pro Arg His Gly Arg Asp Asp Ser Phe Asp Ser Leu Asp Ser
35 40 45

Phe Gly Ser Arg Ser Arg Gln Thr Pro Ser Pro Asp Val Val Leu Arg
50 55 60

Gly Ser Ser Asp Gly Arg Gly Ser Asp Ser Glu Ser Asp Leu Pro His
65 70 75 80

Arg Lys Leu Pro Asp Val Lys Lys Asp Asp Met Ser Ala Arg Arg Thr
85 90 95

Ser His Gly Glu Pro Lys Ser Ala Val Pro Phe Asn Gln Tyr Leu Pro
100 105 110

Asn Lys Ser Asn Gln Thr Ala Tyr Val Pro Ala Pro Leu Arg Lys Lys
115 120 125

Lys Ala Glu Arg Glu Glu Tyr Arg Lys Ser Trp Ser Thr Ala Thr Ser
130 135 140

Pro Leu Gly Gly Glu Arg Pro Phe Ser Phe Pro Glu Thr Ile Glu Glu
145 150 155 160

Glu Gly Ser Glu Val Gly Ser Ala Gly Glu Asp Asn Pro Ala Gly Gln
165 170 175

Met Asn Pro Gly Trp Lys Pro Ser Asp Gly Gly Cys Glu Leu Pro Asp
180 185 190

Gly Ser Gly Lys Glu His Pro Ser Ser Asp Gly Ala Val Val Ala Pro
195 200 205

Ala Pro Lys Ser Glu Glu Lys Asp Ala Ala Glu Ile Gln Lys Arg Lys
210 215 220

Arg Leu Glu Gln Ala Gly Ile Lys Val Met Pro Ala Ala Gln Arg Phe
225 230 235 240

Ala Ser Gln Lys Gln Leu Ser Glu Glu Lys Glu Ala Ile Arg Asp Ile
245 250 255

Val Leu Arg Lys Glu Asn Ser Phe Leu Thr His Gln His Gly Asn Asp
260 265 270

Ser Glu Ala Glu Gly Glu Val Val Cys Arg Leu Pro Asp Leu Glu Lys
275 280 285

Asp Asp Phe Ala Ala Arg Arg Ala Arg Met Asn Gln Thr Lys Pro Met
290 295 300

```
Val Pro Leu Asn Gln Leu Leu Tyr Gly Pro Tyr Pro Lys Lys Gly Ala
305                 310                 315                 320

Glu Lys Ser Asp Gly Ser Lys Gln Leu Ser Lys Gly Ile Ser Lys Lys
                325                 330                 335

Arg Ser Leu Glu Tyr Lys Arg Asn Gln Gly His Thr Glu Glu Val Lys
            340                 345                 350

Leu Ile Val Thr Cys Asn Met Arg Ala Gln Glu Ser Glu Pro Val Glu
        355                 360                 365

Gly Gly Leu Arg Lys Val Pro Asp Leu His Lys Asp Asp Leu Ala Gln
    370                 375                 380

Gln Arg Ile Gln Gly Ser Leu Ala Pro His Arg Glu Pro Pro Ser Phe
385                 390                 395                 400

Ile Thr Leu Ser Asn Ile Thr Glu Ala Asp Leu Glu Thr Trp Glu Arg
                405                 410                 415

Leu Lys Val Ser Glu Lys Ala Arg Asp Gly Asp Val Gln His Ile Cys
            420                 425                 430

Ala Ser Glu Pro Ser Pro Glu Ile Lys Ala Glu Thr Ala Ile Arg Asp
        435                 440                 445

Asp Phe Ala Asn Arg Lys Ala Arg Ala Ser Lys Lys Ala Ser Ser Pro
    450                 455                 460

Arg Gln Lys Phe Val His Phe Gly Pro Val Thr Glu Leu Asp Gln Gln
465                 470                 475                 480

Lys Trp Lys Arg Leu Ser Ile Gly Lys Ala Gly Pro Arg Glu Asp Glu
                485                 490                 495

Glu Glu Val Ile Cys His Gly Ser Lys Ile Gln Met Asp Ser Val Ser
            500                 505                 510

Pro Val Ser Ala Ala Thr Ser Ser Leu Lys Gly His Gln Ile Phe Asn
        515                 520                 525

Arg Gln Asn Asp Cys Arg Thr Met Asn Cys Gly Arg Gly Asp Tyr Cys
    530                 535                 540

Arg Arg Ala Ser Trp Leu Ala Pro Val Pro Glu Ser Gln Glu Glu Trp
545                 550                 555                 560
```

```
Val Cys Ser Leu Gly Glu Cys Pro Arg Gly Thr Glu Glu Val Thr Ser
                565                 570                 575

Lys Gln Leu Pro Gln Asp Gly Lys Glu Glu Thr Glu Ser Ala Pro Arg
            580                 585                 590

Asp Ser Glu Arg Leu Ser Lys Ala Glu Arg Ser Glu Asp Ser Ser Gln
        595                 600                 605

Pro Leu Val Cys Pro Leu Ala Ser Glu Cys Glu Ala Ser Gly Thr Glu
    610                 615                 620

Glu Lys Leu Glu Lys Met Thr Ala Pro Ala Trp Ser Gly Ser Gly Leu
625                 630                 635                 640

Lys Gly Gln Arg Lys Leu Asp Asp Ser Arg Lys Asp Asp Met Met Ala
                645                 650                 655

Arg Arg Thr Gly Met Ser Leu Arg His Thr Gly Ser Asn Pro Asn Gln
            660                 665                 670

Phe Leu Pro Val Pro Phe Ala Lys Gln Gln Asp Val Glu Glu Ser Ser
        675                 680                 685

Lys Gly Leu Pro Met Lys Asp Gln Arg Tyr Gly Pro Arg Thr Pro Val
    690                 695                 700

Ser Asp Asp Ala Glu Ser Thr Ser Met Phe Asp Met Arg Cys Glu Glu
705                 710                 715                 720

Glu Ala Ala Val Gln Pro His Ser Arg Ala Arg Gln Glu Gln Leu Gln
                725                 730                 735

Leu Ile Asn Asn Gln Leu Arg Glu Glu Asp Asp Lys Trp Gln Asp Asp
            740                 745                 750

Leu Ala Arg Trp Lys Ser Arg Arg Arg Ser Val Ser Gln Asp Leu Ile
        755                 760                 765

Lys Lys Glu Glu Glu Arg Lys Lys Met Glu Lys Leu Leu Ala Gly Glu
    770                 775                 780

Asp Gly Thr Ser Glu Arg Arg Lys Ser Ile Lys Thr Tyr Arg Glu Ile
785                 790                 795                 800

Val Gln Glu Lys Glu Arg Arg Glu Arg Glu Leu His Glu Ala Tyr Lys
                805                 810                 815
```

```
Asn Ala Arg Ser Gln Glu Glu Ala Glu Gly Ile Leu Gln Gln Tyr Ile
            820                 825                 830

Glu Arg Phe Thr Ile Ser Glu Ala Val Leu Glu Arg Leu Glu Met Pro
        835                 840                 845

Lys Ile Leu Glu Arg Ser His Ser Thr Glu Pro Asn Leu Ser Ser Phe
    850                 855                 860

Leu Asn Asp Pro Asn Pro Met Lys Tyr Leu Arg Gln Gln Ser Leu Pro
865                 870                 875                 880

Pro Pro Lys Phe Thr Ala Thr Val Glu Thr Thr Ile Ala Arg Ala Ser
                885                 890                 895

Val Leu Asp Thr Ser Met Ser Ala Gly Ser Gly Ser Pro Ser Lys Thr
            900                 905                 910

Val Thr Pro Lys Ala Val Pro Met Leu Thr Pro Lys Pro Tyr Ser Gln
        915                 920                 925

Pro Lys Asn Ser Gln Asp Val Leu Lys Thr Phe Lys Val Asp Gly Lys
    930                 935                 940

Val Ser Val Asn Gly Glu Thr Val His Arg Glu Glu Glu Lys Glu Arg
945                 950                 955                 960

Glu Cys Pro Thr Val Ala Pro Ala His Ser Leu Thr Lys Ser Gln Met
                965                 970                 975

Phe Glu Gly Val Ala Arg Val His Gly Ser Pro Leu Glu Leu Lys Gln
            980                 985                 990

Asp Asn Gly Ser Ile Glu Ile Asn  Ile Lys Lys Pro Asn  Ser Val Pro
        995                 1000                1005

Gln Glu  Leu Ala Ala Thr Thr  Glu Lys Thr Glu Pro  Asn Ser Gln
    1010                1015                1020

Glu Asp  Lys Asn Asp Gly Gly  Lys Ser Arg Lys Gly  Asn Ile Glu
    1025                1030                1035

Leu Ala  Ser Ser Glu Pro Gln  His Phe Thr Thr Thr  Val Thr Arg
    1040                1045                1050

Cys Ser  Pro Thr Val Ala Phe  Val Glu Phe Pro Ser  Ser Pro Gln
    1055                1060                1065
```

```
Leu Lys  Asn Asp Val Ser Glu  Glu Lys Asp Gln Lys  Lys Pro Glu
    1070                 1075                 1080

Asn Glu  Met Ser Gly Lys Val  Glu Leu Val Leu Ser  Gln Lys Val
    1085                 1090                 1095

Val Lys  Pro Lys Ser Pro Glu  Pro Glu Ala Thr Leu  Thr Phe Pro
    1100                 1105                 1110

Phe Leu  Asp Lys Met Pro Glu  Ala Asn Gln Leu His  Leu Pro Asn
    1115                 1120                 1125

Leu Asn  Ser Gln Val Asp Ser  Pro Ser Ser Glu Lys  Ser Pro Val
    1130                 1135                 1140

Met Thr  Pro Phe Lys Phe Trp  Ala Trp Asp Pro Glu  Glu Glu Arg
    1145                 1150                 1155

Arg Arg  Gln Glu Lys Trp Gln  Gln Glu Gln Glu Arg  Leu Leu Gln
    1160                 1165                 1170

Glu Arg  Tyr Gln Lys Glu Gln  Asp Lys Leu Lys Glu  Glu Trp Glu
    1175                 1180                 1185

Lys Ala  Gln Lys Glu Val Glu  Glu Glu Glu Arg Arg  Tyr Tyr Glu
    1190                 1195                 1200

Glu Glu  Arg Lys Ile Ile Glu  Asp Thr Val Val Pro  Phe Thr Val
    1205                 1210                 1215

Ser Ser  Ser Ser Ala Asp Gln  Leu Ser Thr Ser Ser  Ser Met Thr
    1220                 1225                 1230

Glu Gly  Ser Gly Thr Met Asn  Lys Ile Asp Leu Gly  Asn Cys Gln
    1235                 1240                 1245

Asp Glu  Lys Gln Asp Arg Arg  Trp Lys Lys Ser Phe  Gln Gly Asp
    1250                 1255                 1260

Asp Ser  Asp Leu Leu Leu Lys  Thr Arg Glu Ser Asp  Arg Leu Glu
    1265                 1270                 1275

Glu Lys  Gly Ser Leu Thr Glu  Gly Ala Leu Ala His  Ser Gly Asn
    1280                 1285                 1290

Pro Val  Ser Lys Gly Val His  Glu Asp His Gln Leu  Asp Thr Glu
    1295                 1300                 1305
```

```
Ala Gly  Ala Pro His Cys Gly  Thr Asn Pro Gln Leu  Ala Gln Asp
    1310                 1315                 1320

Pro Ser  Gln Asn Gln Gln Thr  Ser Asn Pro Thr His  Ser Ser Glu
    1325                 1330                 1335

Asp Val  Lys Pro Lys Thr Leu  Pro Leu Asp Lys Ser  Ile Asn His
    1340                 1345                 1350

Gln Ile  Glu Ser Pro Ser Glu  Arg Arg Lys Lys Ser  Pro Arg Glu
    1355                 1360                 1365

His Phe  Gln Ala Gly Pro Phe  Ser Pro Cys Ser Pro  Thr Pro Pro
    1370                 1375                 1380

Gly Gln  Ser Pro Asn Arg Ser  Ile Ser Gly Lys Lys  Leu Cys Ser
    1385                 1390                 1395

Ser Cys  Gly Leu Pro Leu Gly  Lys Gly Ala Ala Met  Ile Ile Glu
    1400                 1405                 1410

Thr Leu  Asn Leu Tyr Phe His  Ile Gln Cys Phe Arg  Cys Gly Ile
    1415                 1420                 1425

Cys Lys  Gly Gln Leu Gly Asp  Ala Val Ser Gly Thr  Asp Val Arg
    1430                 1435                 1440

Ile Arg  Asn Gly Leu Leu Asn  Cys Asn Asp Cys Tyr  Met Arg Ser
    1445                 1450                 1455

Arg Ser  Ala Gly Gln Pro Thr  Thr Leu
    I460                 1465
```

<210> 37
<211> 8456
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (3848)..(3848)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (5680)..(5680)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (5735)..(5735)

<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (5750)..(5750)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (5858)..(5858)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (8099)..(8100)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (8108)..(8108)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (8130)..(8130)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (8448)..(8448)
<223> n is a, c, g, or t

<400> 37
ctctcgcggc gcgaccgctc agtcacttcg cccagagacc cggacctggt ccgctgggga 60

gcaggcggcc ataaaccccc tctctcccgg ttccctgacg ccgcggcagg agctgttaca 120

aacaccctgc ggttggtctc cgatgccctt cagtgaggtg gggacgcctg gaccctggtg 180

agcgaacccc aagccacccc ccaccccaac tcagtgtctt cgccggcccc cggcccgtac 240

gcctgtctgg tcgccatggc tgaaaacaca gaggggggatc tgaactccaa cctgctccac 300

gccccctacc acaccgggga ccctcagctg gacacggcca tcgggcagtg gctccgctgg 360

gataagaatc ccaaaacaaa agagcagatt gaaaacctgt tacggaatgg gatgaacaag 420

gagctgcgag atcgtctttg ttgccgaatg acttttggga ctgcaggact tcgttctgcc 480

atgggggcag ggttttgcta tattaatgac cttacagtaa tacagtcaac acaggggatg 540

tacaaatacc ttgagagatg tttctcagac ttcaagcaga gaggctttgt ggttgggtat 600

gacactcggg gtcaagtaac tagcagctgc agcagccaga ggcttgctaa actcactgct 660

gcagtcttgc tggccaaaga tgttcctgtg taccttttt caagatatgt tcctacacct 720

tttgtaccat atgcagttca gaagctcaaa gcagttgcag gtgtgatgat tactgcctct 780

cacaaccgca aggaagacaa tggatacaag gtttactggg aaactggtgc tcagatcaca 840

tctcctcatg ataaagaaat tctaaaatgt atagaagaat gtgtggaacc ctggaatggt 900

tcctggaatg ataatttagt ggataccagc ccgctgaaga gagaccctct gcaggacatt 960
tgcaggagat acatggaaga tctgaaaaag atctgttttt acagggagtt aaactcgaag 1020
accaccttga aatttgtgca cacatctttt catggggtcg gacatgacta tgtgcagttg 1080
gcttttaaag tgtttggttt taagcctcca attccagtac cagaacaaaa agatcctgat 1140
ccagactttt ctaccgttaa atgtccaaat cctgaagaag gagaatctgt gctggaactt 1200
tccttgagac tggcagagaa agaaaatgcc cgggtagtgc tagccacaga tcctgatgca 1260
gacagactgg cagcagcaga acttcaggag aatggttgtt ggaaagtttt cacagggaat 1320
gagttggcag ctttgtttgg atggtggatg tttgattgct ggaagaaaaa taaatcaaga 1380
aatgctgatg tgaagaacgt ttatatgtta gccaccacag tctcttctaa aattctgaag 1440
gcaattgcac ttaaagaagg atttcatttt gaagaaacat taccaggttt taaatggatt 1500
ggaagtagga taatagacct cctggaaaat gggaaagaag tccttttttgc atttgaagag 1560
tctattggtt ttctctgtgg aacttcagtt ttggataaag atggggtgag tgcagctgtt 1620
gtggttgctg agatggcatc ttacctggaa accatgaata taacattgaa acagcaactg 1680
gttaaggttt atgaaaaata tggttatcat atttcaaaaa cttcctattt cttgtgttat 1740
gaaccaccta ccatcaaaag tatatttgaa aggcttcgta attttgattc tccaaaagaa 1800
tatccaaaat tttgtggaac atttgctata ttgcatgtac gggacgttac cactggatat 1860
gacagtagcc agcctaataa gaaatcagtg ctgcctgtga gtaaaaacag ccaaatgatt 1920
acatttactt ttcaaaatgg ctgtgttgct acccttcgga caagtggaac agaaccaaag 1980
ataaagtatt atgcagagat gtgtgcgtca cctgaccaga gtgacactgc tttactggag 2040
gaagaactga agaaactcat tgatgctctg atagagaatt ttcttcagcc tagtaagaat 2100
ggactgatct ggcgttctgt ttaggggtac accaatatgt catgacactg tgtgggcata 2160
tggaacagag caaccgagaa ctgcatacat tgaaccttgt gttagcattc tctctctatc 2220
tcatctggcc gagtatcttt ttccttttat tttctttctt tttggtcaac tggctagact 2280
aaacaaagta taagtttgaa atgaaatggt ctagagggaa atgattcaaa ttttttcata 2340
attcaaacaa aagttattac actaaaagta ttatagtaac gtattgtcct tccgttaaca 2400
gaaacatcac aattcaaaag tgagttttct tattatagtg tgattaagcc tagttctgta 2460
tctcgtaatt gtgtatagca tggtaaaaga aaaaaaaaaa gacaggtaaa tgttatataa 2520
ttgtagggtt agttaaggaa attagtccat aaaattggca agaaaatggt actttccttg 2580
attgttatga aaaagtgtca ttgtgacatc acaggtatta atttacaaat tagagtattt 2640
gaaataggat atttagcatt tctgtagttt agaacagatg ttacagtgtg tggttttggc 2700
atgtctgttt caggtgccct tggtgcagta taaatttact gtatgtattt aaaattctga 2760
tgtttactga acataaaaca atagtggcaa cactcctatt ggctttgttc ttatcaaatt 2820

tctcagtgcc ttctcagggt ctggtatgaa cagaaaaaac agtgttcctg catttaccat 2880
aatttaatca gtcaatttgc agatagctaa ggaaaatctt ccaaaatctg taggtgcttc 2940
ctctatatca agtgagttta ctgattttaa aatcttagga atataaataa agatcttaat 3000
aaaagcaagc ttaatgtagt taaggatctt catattagaa ttgtgtcatt tatcaaacct 3060
tttattacag aaaaaaatct gaactgagaa aggagctctg aatttatttg acagtctaac 3120
ttttgaatta catttttgta agtagattta tcttttattt tatgaagtaa tgttaatttt 3180
actaagaata aaacatccaa attttaagta ttatgaaata ctaaaactac atgttttaaa 3240
acagaggtcc agtttattct aacatcttgc tgacttctgt agaatttctg attgcatgac 3300
tattatttat cttcttttag tgtagaaaat aatttttac aaaaagtcac ttttgggctg 3360
ggcgcagtgg ctcacgcctg taatcccagc actttgggag gctgaggcag gtggatcaca 3420
atgtcaggag tttgagacta gcctggccaa tatggtgaaa ccctgtctct actaaaaata 3480
caaaaattag ccgggtgtgg tggtgggcac ctgtagtccc agctactcgg gaggctgagg 3540
caggagaatt gcttgaacct aggaggctga gattgcagtg agccgaaatc acgccactac 3600
actccagcct gggcaacaga gggagactcc atctcaaaaa aaaaaaaagt cacttttggc 3660
taatcatgca gtcttctgaa aaggtttaat aaaaaaatgc tgtttaggct gggcttggtg 3720
gctcacatct gtaataccag cgctttggga ggctgaggca ggaggattgc ttgaggccag 3780
gagtttaagt tacagaacta tgatcacact gccgcactga agcctggatg acagagtgag 3840
accttatntc aaaaaaaaaa aaaaaaagag agagagagag aaaaagaaat gctgtgttta 3900
atcttcttaa aatgtatctt gtcctacata gttctcagaa tacatatggg gtatagttag 3960
gagataacaa taacaagacc taaacaaatt gcttatggga gatagcactg taattagaga 4020
atttatttca tataaaataa taatataaag tttgaactga ttagaatgtt ccttaaatca 4080
tggtttcgtt tgttgcagct gtcagctaac atttggttga gctcatataa tccaaaccta 4140
attgtccaac tattgtattg ccactattgt taatttgttc cttccttact ttcttattgt 4200
tggatcttat atgactgtat gactgtagat aatagacagc tagctaaatg tagagcacca 4260
tttttctcaa tgagtttgtt ttcaaaggct ttgcactgga gaagatatgt gtagaatttt 4320
tgttttattt ttaaaatgtg taaaacatct tttgttttat atgcttttaa gaatgtagat 4380
tgtaatttag gaaatggcat atattctgca aagtggtata atgttgtata tatgaaaaac 4440
aggtatattt ggtttagttc tgtgtttcag tgactgataa acaaaaatga agctaaaatg 4500
aatcaacggc tcattcatag ttactaaagc catgtcatga ctgtcattct attatgaaga 4560
aaacagtttt atttgagtgc tttaatataa tgcaacattt aagtcatctt aaggtgaaaa 4620
gttattgaag tgttcttctc tcaagttaac aaaattttac aaatgttcct gaagtgtatc 4680
gattgaggga attattttat tctagcctgt catgagtctg tctgagcatg atacttaaca 4740

tgaacttgat gtagtaggtg cacatttgcc taacaggata aattcagtga tagcccaagc 4800
tggtcaggag agacagaagt gctggctagg aggtctcagc ctcgggcact acaatgttta 4860
tctttttctt tttgaacaat attgtgaaaa tgattgcatt ctcattgagc atgaaagtca 4920
aaactggaat tttctaaagc tttggagaaa aggagagaat tagtagggca agatcaaagc 4980
atacattgga aataaacagt agaaataagc aaacaatata aagagacacg ggaaatgctt 5040
ctcgtgaagg gcaaatgtaa caccataaca gtagggaata aactgtgtcc cacctgttta 5100
cttgtgttta tttgttgtta tgatgagtat gcacctacca ttttccttta ttgtgatcac 5160
taaaatgagg ccaaggttta ggaggtgaga aaaagtatcc cttttacgga ttcagttcct 5220
gctaggatgt tgtatcagaa tcatgtgaga actcaggaag aggtagagca gaaataaaaa 5280
ggaagaaaat gtagattaga aaataagact tctgagaaaa agttcaagga ataaagattg 5340
tttgactctg gagaggagaa agctgttgag atagtttgct aagtttgcta ataaagcttt 5400
attgagtctt ttattgaagg gttcaaagta agaaatacta tgccatagtg ttctccatct 5460
ctagtgaaga cagaaccaga gggaataagg atctttttt taaagacagc atggaaagtt 5520
taagatttag gaacttcttg agaaatcttc ataatatctt ctattcaaaa tctttttttt 5580
atttttaagg gaataaaaat tttaaaaatt aagagatata cagatgactt tttaaggtcc 5640
ttctttccag cacatttttt ttaaaaactc aggttagaan aaaaatgtca taaagggcac 5700
taaaagtaac atttgggttt cactccatat ttagnagaat ttagctcacn aaacaatttt 5760
tgtgtgtttt gagtaactga tgcataagtg caatgaataa ttttcctgct tcactgttat 5820
ttttaattta acctggcaaa ttaagtgaga cttgaacnaa attttgttca tcttgaagaa 5880
tgtatagtgg cttaggttat ttgaggtttt cgttttgaat atatatgttg ttattttgat 5940
gtagatacgt tagcagatgg agaggttcta agagattgga tactttggaa tgtctttata 6000
taggaaatag acctcccatt tcttaagcaa gttgcttctg gaattgatgc aacaggggat 6060
attagctgta gtttaatgga taagttgttc tatgagctgc tgctatcagt actttttttcc 6120
acattttcca ttgatatttg tgatgcttga aaaatccaaa cttcacttgg atacatttaa 6180
ttttatgcct ttttctctct acctcccact tttcctaggc acccttgcac attttgttac 6240
tgtgaaggac tattgccctg ttacctcact gataagaaat ctcaactgta ttcattcaga 6300
atgctttaag cttagtcata aactacattt gcctatggac ctttggttat tttgttaatg 6360
ttgtcattat atttaagact tattgtagga agtaatttta aatctatttc ataacctgtt 6420
ttgatctgtc tgcatactgc tatatgttaa tttattacgg cttttacaat tcacttgtca 6480
tctgagattg ttctagaaac tgatgaggaa agagtcgtac taaccaaaat gaatgggcaa 6540
ttgacataac attctatttc aggtttcctt gtacttgggg agccaaagca caagatctaa 6600
ggaaatctta ttcacatttg tccctcagag catcttaccc taagccacag aggtataaat 6660

agtaagatga gaattgctca ttgtctgact ccttgcacag catgaattta attctagttg 6720
gaaacaaagt tagtgttgac tactggttac cttaaagtat tcaccatact atatatttat 6780
tagaccatct ttcctctttg tttattcaag actatgtact tataaaagag aacctataag 6840
aatgtgtaga gtgaccaaat agggacaaag tgacatacct agagtttgct gctcaaattt 6900
ttccactagc gtaaccctaa agcagagggg catgatgaag ggggaagggg ttctgagagc 6960
aaaatcagaa tcagcctacc actgacatga agtgtgttct gtctttaaaa tgtaagtgaa 7020
ttttgcattc tctataataa aacttttaat ataaaaatga tatgtttaat ctttactgaa 7080
acttatgcca ccacaaagga gcaccatgtt catcacgtgg ctgctcagtt ttttccccca 7140
gtgccctcac cgccagtgga ggcaagtcct tgagctacac acactggctg tagtgcaaga 7200
tatataaaca taggccgggc tcggtggctc atgcctgtaa tcctagcact ttgggaggcc 7260
gaagcaggtg gatcatctga ggtcgggaga ttgagactag cctgaccgac atggagaaac 7320
cctgtgtcca ctaaaaatac aaaattagcc gggtgtggtg gcacatgcct gtaatcccag 7380
ctactcggga ggctgaggca ggagaatcac ttgaacccag gaggtggagg ttgtggtgag 7440
ccaagatcgc gccgttgcac tccagcctgg gcaacaagag tgaaactccg tctcaaaaaa 7500
aaaaaaaaaa gatatataaa catagaacat atccagacta tttcctgccc ttggaatctg 7560
accttcagaa ggattaacgc caaactaaac tgacctgaaa cagttttgtt actattggca 7620
ttccttgact aaaaaggacc acctgatgct cagtcccgag agtaggaatc gtggtttatt 7680
atttttcata tccataccac ctagttcagt acctgttata gagtgtttaa ttgctgatac 7740
taggaattaa gggatttttt aaaagtagca aatagtttaa ggaacaatca aaaaatgaat 7800
tttctaattc agtttttca atgaagtccc atcactgaac tttataatgt aatgtttatt 7860
aagcagacat ccccttaaat actcagtttc tttggagaga gcaaagcata tctccctgtg 7920
gaaattatat ataatgcctg gtattttaca ctcaaagcaa ggtgttggcc aggaaggtaa 7980
agaaatggtg tctatgtaga cttaatatca ctagtctaag tgtcattttg agcctagtag 8040
cactaccttc caagtgagtc acaacaaatt tgatcctatt tgtaggtttt ttccactgnn 8100
atgattcntc atgtatctta caagagccan tcaagcaaga ctctgcttct atgtatggtg 8160
aggccttgtt gttctagggt agaataaact ctttgtatgc ctcattgaat atgccaggta 8220
aaatatatgc agtcaagaat gaattatttt tctgactaaa gtgtgtagca gtagttcaaa 8280
attgtgccct tgttttaaca gtttctgtca acatcttctc atttttccct acaaaaacac 8340
cagggtgtat tataagtact gcctgtgaga atttgcactt tatgtatttg tgtgtggatt 8400
tcttgtggtt ttagccaaat gaagtgttat cagtaataaa caggtctntt catagg 8456

<210> 38
<211> 622

<212> PRT
<213> Homo sapiens

<400> 38

Met Ala Glu Asn Thr Glu Gly Asp Leu Asn Ser Asn Leu Leu His Ala
1 5 10 15

Pro Tyr His Thr Gly Asp Pro Gln Leu Asp Thr Ala Ile Gly Gln Trp
20 25 30

Leu Arg Trp Asp Lys Asn Pro Lys Thr Lys Glu Gln Ile Glu Asn Leu
35 40 45

Leu Arg Asn Gly Met Asn Lys Glu Leu Arg Asp Arg Leu Cys Cys Arg
50 55 60

Met Thr Phe Gly Thr Ala Gly Leu Arg Ser Ala Met Gly Ala Gly Phe
65 70 75 80

Cys Tyr Ile Asn Asp Leu Thr Val Ile Gln Ser Thr Gln Gly Met Tyr
85 90 95

Lys Tyr Leu Glu Arg Cys Phe Ser Asp Phe Lys Gln Arg Gly Phe Val
100 105 110

Val Gly Tyr Asp Thr Arg Gly Gln Val Thr Ser Ser Cys Ser Ser Gln
115 120 125

Arg Leu Ala Lys Leu Thr Ala Ala Val Leu Leu Ala Lys Asp Val Pro
130 135 140

Val Tyr Leu Phe Ser Arg Tyr Val Pro Thr Pro Phe Val Pro Tyr Ala
145 150 155 160

Val Gln Lys Leu Lys Ala Val Ala Gly Val Met Ile Thr Ala Ser His
165 170 175

Asn Arg Lys Glu Asp Asn Gly Tyr Lys Val Tyr Trp Glu Thr Gly Ala
180 185 190

Gln Ile Thr Ser Pro His Asp Lys Glu Ile Leu Lys Cys Ile Glu Glu
195 200 205

Cys Val Glu Pro Trp Asn Gly Ser Trp Asn Asp Asn Leu Val Asp Thr
210 215 220

Ser Pro Leu Lys Arg Asp Pro Leu Gln Asp Ile Cys Arg Arg Tyr Met
225 230 235 240

```
Glu Asp Leu Lys Lys Ile Cys Phe Tyr Arg Glu Leu Asn Ser Lys Thr
                245                 250                 255

Thr Leu Lys Phe Val His Thr Ser Phe His Gly Val Gly His Asp Tyr
            260                 265                 270

Val Gln Leu Ala Phe Lys Val Phe Gly Phe Lys Pro Pro Ile Pro Val
        275                 280                 285

Pro Glu Gln Lys Asp Pro Asp Pro Asp Phe Ser Thr Val Lys Cys Pro
    290                 295                 300

Asn Pro Glu Glu Gly Glu Ser Val Leu Glu Leu Ser Leu Arg Leu Ala
305                 310                 315                 320

Glu Lys Glu Asn Ala Arg Val Val Leu Ala Thr Asp Pro Asp Ala Asp
                325                 330                 335

Arg Leu Ala Ala Ala Glu Leu Gln Glu Asn Gly Cys Trp Lys Val Phe
            340                 345                 350

Thr Gly Asn Glu Leu Ala Ala Leu Phe Gly Trp Trp Met Phe Asp Cys
        355                 360                 365

Trp Lys Lys Asn Lys Ser Arg Asn Ala Asp Val Lys Asn Val Tyr Met
    370                 375                 380

Leu Ala Thr Thr Val Ser Ser Lys Ile Leu Lys Ala Ile Ala Leu Lys
385                 390                 395                 400

Glu Gly Phe His Phe Glu Glu Thr Leu Pro Gly Phe Lys Trp Ile Gly
                405                 410                 415

Ser Arg Ile Ile Asp Leu Leu Glu Asn Gly Lys Glu Val Leu Phe Ala
            420                 425                 430

Phe Glu Glu Ser Ile Gly Phe Leu Cys Gly Thr Ser Val Leu Asp Lys
        435                 440                 445

Asp Gly Val Ser Ala Ala Val Val Val Ala Glu Met Ala Ser Tyr Leu
    450                 455                 460

Glu Thr Met Asn Ile Thr Leu Lys Gln Gln Leu Val Lys Val Tyr Glu
465                 470                 475                 480

Lys Tyr Gly Tyr His Ile Ser Lys Thr Ser Tyr Phe Leu Cys Tyr Glu
                485                 490                 495
```

Pro Pro Thr Ile Lys Ser Ile Phe Glu Arg Leu Arg Asn Phe Asp Ser
500 505 510

Pro Lys Glu Tyr Pro Lys Phe Cys Gly Thr Phe Ala Ile Leu His Val
515 520 525

Arg Asp Val Thr Thr Gly Tyr Asp Ser Ser Gln Pro Asn Lys Lys Ser
530 535 540

Val Leu Pro Val Ser Lys Asn Ser Gln Met Ile Thr Phe Thr Phe Gln
545 550 555 560

Asn Gly Cys Val Ala Thr Leu Arg Thr Ser Gly Thr Glu Pro Lys Ile
565 570 575

Lys Tyr Tyr Ala Glu Met Cys Ala Ser Pro Asp Gln Ser Asp Thr Ala
580 585 590

Leu Leu Glu Glu Glu Leu Lys Lys Leu Ile Asp Ala Leu Ile Glu Asn
595 600 605

Phe Leu Gln Pro Ser Lys Asn Gly Leu Ile Trp Arg Ser Val
610 615 620

<210> 39
<211> 2015
<212> DNA
<213> Homo sapiens

<400> 39
acaaataata aggaacaagg tgtggttatt gattctctaa aattaagtga ggagatgaag 60

cccaatctag atggtgttga tttattcaac aatggtggtt ctggaaacgg tgaaacgaaa 120

actgggctga gactgaaagc aataaatctg cctttggaaa atgaagtaac tgaaatttca 180

gctttgcagg tgcatttgga tgaattccaa aaaatcttat ggaaggaaag agaaatgcgc 240

acagctttgg aaaaagaaat agagagactg gagtcggctt tgtctctgtg gaagtggaag 300

tatgaagaac tgaaagaatc aaagccaaaa aatgtgaaag agtttgacat tcttcttggt 360

caacataatg atgaaatgca agaactgtca ggcaatataa aggaagaatc caaatctcaa 420

aacagcaaag acagagtgat ttgtgagtta agagcagagc tagagagatt gcaagctgaa 480

aatacctcgg agtgggacaa gagggaaata cttgaaagag aaaagcaggg actggagaga 540

gaaaatagaa ggctgaagat ccaggtgaaa gaaatggaag agcttttgga taagaaaaat 600

agattaagtg caaactctca aagtcctgat ttcaagatgt cacaaattga tctgcaagaa 660

aaaaaccagg aattactgaa ccttcaacat gcctactata aactaaacag acaataccag 720

gcaaatattg cagaactgac tcatgcaaac aaccgagtgg atcaaaatga agcagaagta 780

aagaaactaa gattacgagt ggaagaacta aagcagggac tcaaccaaaa agaagatgag 840

cttgatgatt ccctgaatca gatccgtaag ctccagaggt ctctggatga agagaaagaa 900

agaaatgaaa acttagagac tgaactcagg cacttgcaaa actggtaatt ttttcacaaa 960

atatgctgaa ttaaagatta gggccttaaa gacatttcca tatccttttc ttaaatatca 1020

gtaaaattgt ttttattaac tagaaatatt aatgaaaaaa acgtagacaa tacacaaatt 1080

aatgggcttc ttcacttctt ctaatttttg cctaacagat actgcatatt ctcaaaaaga 1140

caatttaaat gtcatttaaa aacaacttta attctaagat gtgtaaatat tttgaaagtc 1200

aaaaagggct ttcagaatac tttttacata aaatctgaaa gagttataat atcggtaaga 1260

aaaagtaagt tgaaaaccat acaagacgct gggtcattaa taagaaaacc attgacttta 1320

agtataaagt actggtttgt ttaaataatt ggtaaacttt tatgtacgtg ttgtctatgt 1380

ggtggggatg gcaggttgta ttaacaaaaa tgaatcattc tagaggtgta acaatacatt 1440

tcttatataa ttttataagt catttctaat ctttgtataa aacagaagtg agcaagatga 1500

atcagaaaaa aggtgttttg tattttaaag gtaacagata accaggtgat tgaatctaag 1560

acaggctgta agcatcgctg agaaactaaa aggacttttg acttttatct ggatagacat 1620

ttctacagta aaatcatgga aaggcatcag cattgcaaag tagcatctag gtagaaatca 1680

ggccaaaatt aagctgtggt ttccctctga gtagtgggaa tagagaaaat taggaaattg 1740

tggttatgtg aatatttctt taaaactttt atgtacatta tagtttattg cttcatattt 1800

aagtttagtt tttaaggtaa aatgttattt tgaacaaaaa gacacttata attttccata 1860

cctattttca actgaaggca acttgtaaga tttaactcag tcaataacat actggtttta 1920

ctcatctccc cctccattga ttagccaaaa aaaaatgaaa tcttactaat tcattattga 1980

ataaagacca cttttatcag agtcaaaaaa aaaaa 2015

<210> 40
<211> 297
<212> PRT
<213> Homo sapiens

<400> 40

Met Lys Pro Asn Leu Asp Gly Val Asp Leu Phe Asn Asn Gly Gly Ser
1 5 10 15

Gly Asn Gly Glu Thr Lys Thr Gly Leu Arg Leu Lys Ala Ile Asn Leu
20 25 30

Pro Leu Glu Asn Glu Val Thr Glu Ile Ser Ala Leu Gln Val His Leu
35 40 45

Asp Glu Phe Gln Lys Ile Leu Trp Lys Glu Arg Glu Met Arg Thr Ala
50 55 60

Leu Glu Lys Glu Ile Glu Arg Leu Glu Ser Ala Leu Ser Leu Trp Lys
65 70 75 80

Trp Lys Tyr Glu Glu Leu Lys Glu Ser Lys Pro Lys Asn Val Lys Glu
85 90 95

Phe Asp Ile Leu Leu Gly Gln His Asn Asp Glu Met Gln Glu Leu Ser
100 105 110

Gly Asn Ile Lys Glu Glu Ser Lys Ser Gln Asn Ser Lys Asp Arg Val
115 120 125

Ile Cys Glu Leu Arg Ala Glu Leu Glu Arg Leu Gln Ala Glu Asn Thr
130 135 140

Ser Glu Trp Asp Lys Arg Glu Ile Leu Glu Arg Glu Lys Gln Gly Leu
145 150 155 160

Glu Arg Glu Asn Arg Arg Leu Lys Ile Gln Val Lys Glu Met Glu Glu
165 170 175

Leu Leu Asp Lys Lys Asn Arg Leu Ser Ala Asn Ser Gln Ser Pro Asp
180 185 190

Phe Lys Met Ser Gln Ile Asp Leu Gln Glu Lys Asn Gln Glu Leu Leu
195 200 205

Asn Leu Gln His Ala Tyr Tyr Lys Leu Asn Arg Gln Tyr Gln Ala Asn
210 215 220

Ile Ala Glu Leu Thr His Ala Asn Asn Arg Val Asp Gln Asn Glu Ala
225 230 235 240

Glu Val Lys Lys Leu Arg Leu Arg Val Glu Glu Leu Lys Gln Gly Leu
245 250 255

Asn Gln Lys Glu Asp Glu Leu Asp Asp Ser Leu Asn Gln Ile Arg Lys
260 265 270

Leu Gln Arg Ser Leu Asp Glu Glu Lys Glu Arg Asn Glu Asn Leu Glu
275 280 285

Thr Glu Leu Arg His Leu Gln Asn Trp
290 295

<210> 41
<211> 2372
<212> DNA
<213> Hmo sapiens

<400> 41
cgataatgac tttgcacgat tcactttggg atctcaaagt gcttccaaag cattcagatt 60

tacaaacaat tcacaagaca ggtcatcttt gtaataccca tacttacaac gaattaacaa 120

gaggagtgac ttaagattct ccaggaacac agtggcagct attgatgatc tgttttctat 180

ctgtttgata gagcatcatg agaaatcaca aaatacaatg ctatttttct gatgtgtgct 240

aataaagtca aagaaagcaa atacatcttg acacttttgt ccattttcat taaaaaaaaa 300

aaaaagttca gggtgtttgg aattttacac ctcagcacac cttactggta tcaatggata 360

aagcgggtaa ttgacagatc cacccaaatg ccactgcagt cagaagcaga tctggacaca 420

cccttgttta cagtttcata ttgggttgct atagttcccg tgctaaatca ccagctttca 480

ggaacatgac tgctcctggc agtggaaggt gctgaaacag aaattttaat taaaaacttt 540

atcaagtact cttcacagtg ctgcttggca ccatagaaaa tcagtacaat atatcgagcc 600

ctactttgga ggagctggat ttctgaggga gctgatccag ttctaagtgt cttctcgaat 660

taggagatag atgatctttg atggggatct cctccgtcac cacaggccag tcacagaacc 720

aactagccac gtgctgccag acctcagtgg gcccaagcag gagcaatctc ttctatcccc 780

catctccccc aggaccatcc cccccattgt caatgtcatc cagggctctt ctggtagtga 840

gtgacttttc tgcacatgtt tagggcttgg gggagctaga acacaggaaa catgaatgca 900

aaaggcatgg aaaacactgt tttgctttgg gttagtaaaa tgtgggcagg acaaagatta 960

ctattggtcg gagctttgcc aagtgagata gaatcaactg tcaccccatt cctttcccag 1020

aaggtcttat ggtattaagg atacatccag tattttccca cagattttta ttcaggcgat 1080

gtttcataaa ttacatatat gaaaacattc attattacat ttccttgtgt gtttcaaaca 1140

gacattggca ccttcctatt gagttaattc tctgcatctt ttgcagcagc agcccgcaag 1200

gagattccca gagatggctc ccctaacaca cagtcctgtg attttacagt tctatgactt 1260

acagttgatg attcacaaga ttcaggattc tacaagactc aaggggggaac taaactttct 1320

tacgattgta catgatcagt tatagggctg taatcattaa ttgttggctt caaatgtgga 1380

cacacacaca cacacatcat gccaaggagg gaatggggtg tttcaagtca ggcagcgatg 1440

attctggaag gttggaaatg taaggttaga agcttggctg gtcttagtaa acttgttccc 1500

ttgctcccac caagaagagg taccaaatgt gagacctgag atctcctcca atatctgtcc 1560

tctgcagttc cgggaaacta atcatgaagt acacatgcag cagctcctcc acttcctttc 1620

ctccgaggtc ctcctttcca ttctcccacc tagatactga cacaccgcca cggtttccac 1680

attggaaggg cagaacactg tgcagtatcg tgcacacttg ctgggttagg aatagagctg 1740

ccctagggtc accttcatgc aagtattgac agctacaaat taaagtcctt agagcagttg 1800

acacagatac tacgttctag aagagaatta aatttaaacg tcaagtttaa agggatcata 1860

attctgcagg tatctttctc tgagtgactg aatgtgacta ttgcattagg gtaaatgaat 1920

taagacgtgc aagtgggatt tactgtatgt tagaaaggag ttttgcagcc aagactgcct 1980

tgaataaaat gtgtttgcac tgaaaaaaaa ttttaaatta cttggtctct ggttgctgta 2040

aaggtcatcc aagatggatg ttctgtttat attgtatagt atttcatatg aaataattac 2100

agttcatgaa atgtcttccc taatgttact gatttataac agcacatttg taacatggtt 2160

tttatcgtgt cagtgtacca tactgtaaat gatgattact tgtcatgctt agtataataa 2220

cttaaaagaa aaaaaaggac agggattttt gtaagtctat atttgaaagt ccctccctat 2280

ggtgatactg tgttcatgtt gtttatgtag tgttgtgtga aatatccatt ttggattgtg 2340

ttacttttta agatattaaa taacatttgg tt 2372

<210> 42
<211> 365
<212> PRT
<213> Homo sapiens

<400> 42

Lys Arg Lys Val Ala Gln Glu Ala Ile Glu Ser Leu Ser Ala Ser Lys
1 5 10 15

Leu Ala Lys Ser Ile Cys Ser Gln Phe Arg Thr Arg Leu Asn Ser Ser
20 25 30

His Glu Ala Phe Ala Ala Ser Leu Arg Gln Leu Glu Ala Gly His Ser
35 40 45

Gly Arg Leu Glu Lys Thr Glu Asp Leu Trp Leu Arg Val Arg Lys Asp
50 55 60

His Ala Pro Arg Leu Ala Arg Leu Ser Leu Glu Ser Arg Ser Leu Gln
65 70 75 80

Asp Val Leu Leu His Arg Lys Pro Lys Leu Gly Gln Glu Leu Gly Arg
85 90 95

Gly Gln Tyr Gly Val Val Tyr Leu Cys Asp Asn Trp Gly Gly His Phe
100 105 110

Pro Cys Ala Leu Lys Ser Val Val Pro Pro Asp Glu Lys His Trp Asn
115 120 125

Asp Leu Ala Leu Glu Phe His Tyr Met Arg Ser Leu Pro Lys His Glu

130 135 140

Arg Leu Val Asp Leu His Gly Ser Val Ile Asp Tyr Asn Tyr Gly Gly
145 150 155 160

Gly Ser Ser Ile Ala Val Leu Leu Ile Met Glu Arg Leu His Arg Asp
165 170 175

Leu Tyr Thr Gly Leu Lys Ala Gly Leu Thr Leu Glu Thr Arg Leu Gln
180 185 190

Ile Ala Leu Asp Val Val Glu Gly Ile Arg Phe Leu His Ser Gln Gly
195 200 205

Leu Val His Arg Asp Ile Lys Leu Lys Asn Val Leu Leu Asp Lys Gln
210 215 220

Asn Arg Ala Lys Ile Thr Asp Leu Gly Phe Cys Lys Pro Glu Ala Met
225 230 235 240

Met Ser Gly Ser Ile Val Gly Thr Pro Ile His Met Ala Pro Glu Leu
245 250 255

Phe Thr Gly Lys Tyr Asp Asn Ser Val Asp Val Tyr Ala Phe Gly Ile
260 265 270

Leu Phe Trp Tyr Ile Cys Ser Gly Ser Val Lys Leu Pro Glu Ala Phe
275 280 285

Glu Arg Cys Ala Ser Lys Asp His Leu Trp Asn Asn Val Arg Arg Gly
290 295 300

Ala Arg Pro Glu Arg Leu Pro Val Phe Asp Glu Glu Cys Trp Gln Leu
305 310 315 320

Met Glu Ala Cys Trp Asp Gly Asp Pro Leu Lys Arg Pro Leu Leu Gly
325 330 335

Ile Val Gln Pro Met Leu Gln Gly Ile Met Asn Arg Leu Cys Lys Ser
340 345 350

Asn Ser Glu Gln Pro Asn Arg Gly Leu Asp Asp Ser Thr
355 360 365

<210> 43
<211> 1185
<212> PRT
<213> Homo sapiens

<400> 43

Lys Lys Ile Arg Glu Lys Phe Asn Arg Tyr Leu Asp Val Val Asn Arg
1 5 10 15

Asn Lys Gln Val Val Glu Ala Ser Tyr Thr Ala His Leu Thr Ser Pro
20 25 30

Leu Thr Ala Ile Gln Asp Cys Cys Thr Ile Pro Pro Ser Met Met Glu
35 40 45

Phe Asp Gly Asn Phe Asn Thr Asn Val Ser Arg Thr Ile Ser Cys Asp
50 55 60

Arg Leu Ser Thr Thr Val Asn Ser Arg Ala Phe Asn Pro Gly Arg Asp
65 70 75 80

Leu Asn Ser Val Leu Ala Asp Asn Leu Lys Ser Asn Pro Gly Ile Lys
85 90 95

Trp Gln Tyr Phe Ser Ser Glu Glu Gly Ile Phe Thr Val Phe Pro Ala
100 105 110

His Lys Phe Arg Cys Lys Gly Ser Tyr Glu His Arg Ser Arg Pro Ile
115 120 125

Tyr Val Ser Thr Val Arg Pro Gln Ser Lys His Ile Val Val Ile Leu
130 135 140

Asp His Gly Ala Ser Val Thr Asp Thr Gln Leu Gln Ile Ala Lys Asp
145 150 155 160

Ala Ala Gln Val Ile Leu Ser Ala Ile Asp Glu His Asp Lys Ile Ser
165 170 175

Val Leu Thr Val Ala Asp Thr Val Arg Thr Cys Ser Leu Asp Gln Cys
180 185 190

Tyr Lys Thr Phe Leu Ser Pro Ala Thr Ser Glu Thr Lys Arg Lys Met
195 200 205

Ser Thr Phe Val Ser Ser Val Lys Ser Ser Asp Ser Pro Thr Gln His
210 215 220

Ala Val Gly Phe Gln Lys Ala Phe Gln Leu Ile Arg Ser Thr Asn Asn
225 230 235 240

Asn Thr Lys Phe Gln Ala Asn Thr Asp Met Val Ile Ile Tyr Leu Ser

245 250 255

Ala Gly Ile Thr Ser Lys Asp Ser Ser Glu Glu Asp Lys Lys Ala Thr
260 265 270

Leu Gln Val Ile Asn Glu Glu Asn Ser Phe Leu Asn Asn Ser Val Met
275 280 285

Ile Leu Thr Tyr Ala Leu Met Asn Asp Gly Val Thr Gly Leu Lys Glu
290 295 300

Leu Ala Phe Leu Arg Asp Leu Ala Glu Gln Asn Ser Gly Lys Tyr Gly
305 310 315 320

Val Pro Asp Arg Thr Ala Leu Pro Val Ile Lys Gly Ser Met Met Val
325 330 335

Leu Asn Gln Leu Ser Asn Leu Glu Thr Thr Val Gly Arg Phe Tyr Thr
340 345 350

Asn Leu Pro Asn Arg Met Ile Asp Glu Ala Val Phe Ser Leu Pro Phe
355 360 365

Ser Asp Glu Met Gly Asp Gly Leu Ile Met Thr Val Ser Lys Pro Cys
370 375 380

Tyr Phe Gly Asn Leu Leu Leu Gly Ile Val Gly Val Asp Val Asn Leu
385 390 395 400

Ala Tyr Ile Leu Glu Asp Val Thr Tyr Tyr Gln Asp Ser Leu Ala Ser
405 410 415

Tyr Thr Phe Leu Ile Asp Asp Lys Gly Tyr Thr Leu Met His Pro Ser
420 425 430

Leu Thr Arg Pro Tyr Leu Leu Ser Glu Pro Pro Leu His Thr Asp Ile
435 440 445

Ile His Tyr Glu Asn Ile Pro Lys Phe Glu Leu Val Arg Gln Asn Ile
450 455 460

Leu Ser Leu Pro Leu Gly Ser Gln Ile Ile Ala Val Pro Val Asn Ser
465 470 475 480

Ser Leu Ser Trp His Ile Asn Lys Leu Arg Glu Thr Gly Lys Glu Ala
485 490 495

Tyr Asn Val Ser Tyr Ala Trp Lys Met Val Gln Asp Thr Ser Phe Ile
500 505 510

Leu Cys Ile Val Val Ile Gln Pro Glu Ile Pro Val Lys Gln Leu Lys
515 520 525

Asn Leu Asn Thr Val Pro Ser Ser Lys Leu Leu Tyr His Arg Leu Asp
530 535 540

Leu Leu Gly Gln Pro Ser Ala Cys Leu His Phe Lys Gln Leu Ala Thr
545 550 555 560

Leu Glu Ser Pro Thr Ile Met Leu Ser Ala Gly Ser Phe Ser Ser Pro
565 570 575

Tyr Glu His Leu Ser Gln Pro Glu Thr Lys Arg Met Val Glu His Tyr
580 585 590

Thr Ala Tyr Leu Ser Asp Asn Thr Arg Leu Ile Ala Asn Pro Gly Leu
595 600 605

Lys Phe Ser Val Arg Asn Glu Val Met Ala Thr Ser His Val Thr Asp
610 615 620

Glu Trp Met Thr Gln Met Glu Met Ser Ser Leu Asn Thr Tyr Ile Val
625 630 635 640

Arg Arg Tyr Ile Ala Thr Pro Asn Gly Val Leu Arg Ile Tyr Pro Gly
645 650 655

Ser Leu Met Asp Lys Ala Phe Asp Pro Thr Arg Arg Gln Trp Tyr Leu
660 665 670

His Ala Val Ala Asn Pro Gly Leu Ile Ser Leu Thr Gly Pro Tyr Leu
675 680 685

Asp Val Gly Gly Ala Gly Tyr Val Val Thr Ile Ser His Thr Ile His
690 695 700

Ser Ser Ser Thr Gln Leu Ser Ser Gly His Thr Val Ala Val Met Gly
705 710 715 720

Ile Asp Phe Thr Leu Arg Tyr Phe Tyr Lys Val Leu Met Asp Leu Leu
725 730 735

Pro Val Cys Asn Gln Asp Gly Gly Asn Lys Ile Arg Cys Phe Ile Met
740 745 750

```
Glu Asp Arg Gly Tyr Leu Val Ala His Pro Thr Leu Ile Asp Pro Lys
        755                 760                 765

Gly His Ala Pro Val Glu Gln Gln His Ile Thr His Lys Glu Pro Leu
    770                 775                 780

Val Ala Asn Asp Ile Leu Asn His Pro Asn Phe Val Lys Lys Asn Leu
785                 790                 795                 800

Cys Asn Ser Phe Ser Asp Arg Thr Val Gln Arg Phe Tyr Lys Phe Asn
                805                 810                 815

Thr Ser Leu Ala Gly Asp Leu Thr Asn Leu Val His Gly Ser His Cys
            820                 825                 830

Ser Lys Tyr Arg Leu Ala Arg Ile Pro Gly Thr Asn Ala Phe Val Gly
        835                 840                 845

Ile Val Asn Glu Thr Cys Asp Ser Leu Ala Phe Cys Ala Cys Ser Met
    850                 855                 860

Val Asp Arg Leu Cys Leu Asn Cys His Arg Met Glu Gln Asn Glu Cys
865                 870                 875                 880

Glu Cys Pro Cys Glu Cys Pro Leu Glu Val Asn Glu Cys Thr Gly Asn
                885                 890                 895

Leu Thr Asn Ala Glu Asn Arg Asn Pro Ser Cys Glu Val His Gln Glu
            900                 905                 910

Pro Val Thr Tyr Thr Ala Ile Asp Pro Gly Leu Gln Asp Ala Leu His
        915                 920                 925

Gln Cys Val Asn Ser Arg Cys Ser Gln Arg Leu Glu Ser Gly Asp Cys
    930                 935                 940

Phe Gly Val Leu Asp Cys Glu Trp Cys Met Val Asp Ser Asp Gly Lys
945                 950                 955                 960

Thr His Leu Asp Lys Pro Tyr Cys Ala Pro Gln Lys Glu Cys Phe Gly
                965                 970                 975

Gly Ile Val Gly Ala Lys Ser Pro Tyr Val Asp Asp Met Gly Ala Ile
            980                 985                 990

Gly Asp Glu Val Ile Thr Leu Asn  Met Ile Lys Ser Ala  Pro Val Gly
        995                 1000                1005
```

```
Pro Val  Ala Gly Gly Ile Met  Gly Cys Ile Met Val  Leu Val Leu
    1010                1015                1020

Ala Val  Tyr Ala Tyr Arg His  Gln Ile His Arg Arg  Ser His Gln
    1025                1030                1035

His Met  Ser Pro Leu Ala Ala  Gln Glu Met Ser Val  Arg Met Ser
    1040                1045                1050

Asn Leu  Glu Asn Asp Arg Asp  Glu Arg Asp Asp Asp  Ser His Glu
    1055                1060                1065

Asp Arg  Gly Ile Ile Ser Asn  Thr Arg Phe Ile Ala  Ala Val Ile
    1070                1075                1080

Glu Arg  His Ala His Ser Pro  Glu Arg Arg Arg Arg  Tyr Trp Gly
    1085                1090                1095

Arg Ser  Gly Thr Glu Ser Asp  His Gly Tyr Ser Thr  Met Ser Pro
    1100                1105                1110

Gln Glu  Asp Ser Glu Asn Pro  Pro Cys Asn Asn Asp  Pro Leu Ser
    1115                1120                1125

Ala Gly  Val Asp Val Gly Asn  His Asp Glu Asp Leu  Asp Leu Asp
    1130                1135                1140

Thr Pro  Pro Gln Thr Ala Ala  Leu Leu Ser His Lys  Phe His His
    1145                1150                1155

Tyr Arg  Ser His His Pro Thr  Leu His His Ser His  His Leu Gln
    1160                1165                1170

Ala Ala  Val Thr Val His Thr  Val Asp Ala Glu Cys
    1175                1180                1185
```

<210> 44
<211> 410
<212> PRT
<213> Homo sapiens

<400> 44

```
Met Asp Glu Val Ser Leu Leu Ser Glu Leu Val Glu Ala Ala Ser Phe
1               5                  10                  15

Leu Gln Val Thr Ser Leu Leu Gln Leu Leu Leu Ser Gln Val Arg Leu
            20                  25                  30
```

```
Asn Asn Cys Leu Glu Met Tyr Arg Leu Ala Gln Val Tyr Gly Leu Pro
        35                  40                  45

Asp Leu Gln Glu Ala Cys Leu Arg Phe Met Val Val His Phe His Glu
    50                  55                  60

Val Leu Cys Lys Pro Gln Phe His Leu Leu Gly Ser Pro Pro Gln Ala
65                  70                  75                  80

Pro Gly Asp Val Ser Leu Lys Gln Arg Leu Arg Glu Ala Arg Met Thr
                85                  90                  95

Gly Thr Pro Val Leu Val Ala Leu Gly Asp Phe Leu Gly Gly Pro Leu
            100                 105                 110

Ala Pro His Pro Tyr Gln Gly Glu Pro Pro Ser Met Leu Arg Tyr Glu
        115                 120                 125

Glu Met Thr Glu Arg Trp Phe Pro Leu Ala Asn Asn Leu Pro Pro Asp
    130                 135                 140

Leu Val Asn Val Arg Gly Tyr Gly Ser Ala Ile Leu Asp Asn Tyr Leu
145                 150                 155                 160

Phe Ile Val Gly Gly Tyr Arg Ile Thr Ser Gln Glu Ile Ser Ala Ala
                165                 170                 175

His Ser Tyr Asn Pro Ser Thr Asn Glu Trp Leu Gln Val Ala Ser Met
            180                 185                 190

Asn Gln Lys Arg Ser Asn Phe Lys Leu Val Ala Val Asn Ser Lys Leu
        195                 200                 205

Tyr Ala Ile Gly Gly Gln Ala Val Ser Asn Val Glu Cys Tyr Asn Pro
    210                 215                 220

Glu Gln Asp Ala Trp Asn Phe Val Ala Pro Leu Pro Asn Pro Leu Ala
225                 230                 235                 240

Glu Phe Ser Ala Cys Glu Cys Lys Gly Lys Ile Tyr Val Ile Gly Gly
                245                 250                 255

Tyr Thr Thr Arg Asp Arg Asn Met Asn Ile Leu Gln Tyr Cys Pro Ser
            260                 265                 270

Ser Asp Met Trp Thr Leu Phe Glu Thr Cys Asp Val His Ile Arg Lys
        275                 280                 285
```

```
Gln Gln Met Val Ser Val Glu Glu Thr Ile Tyr Ile Val Gly Gly Cys
    290                 295                 300

Leu His Glu Leu Gly Pro Asn Arg Arg Ser Ser Gln Ser Glu Asp Met
305                 310                 315                 320

Leu Thr Val Gln Ser Tyr Asn Thr Val Thr Arg Gln Trp Leu Tyr Leu
                325                 330                 335

Lys Glu Asn Thr Ser Lys Ser Gly Leu Asn Leu Thr Cys Ala Leu His
            340                 345                 350

Asn Asp Gly Ile Tyr Ile Met Ser Arg Asp Val Thr Leu Ser Thr Ser
        355                 360                 365

Leu Glu His Arg Val Phe Leu Lys Tyr Asn Ile Phe Ser Asp Ser Trp
    370                 375                 380

Glu Ala Phe Arg Arg Phe Pro Ala Phe Gly His Asn Leu Leu Val Ser
385                 390                 395                 400

Ser Leu Tyr Leu Pro Asn Lys Ala Glu Thr
                405                 410
```

<210> 45
<211> 1101
<212> PRT
<213> Homo sapiens

<400> 45

```
Arg Arg Arg Arg Pro Ser Ser Ser Arg Arg Leu Arg Gly Arg Gly Ala
1               5                   10                  15

Ala Gln Met Ala Cys Pro Ala Leu Gly Leu Glu Ala Leu Gln Pro Leu
            20                  25                  30

Gln Pro Glu Pro Pro Pro Glu Pro Ala Phe Ser Glu Ala Gln Lys Trp
        35                  40                  45

Ile Glu Gln Val Thr Gly Arg Ser Phe Gly Asp Lys Asp Phe Arg Thr
    50                  55                  60

Gly Leu Glu Asn Gly Ile Leu Leu Cys Glu Leu Leu Asn Ala Ile Lys
65                  70                  75                  80

Pro Gly Leu Val Lys Lys Ile Asn Arg Leu Pro Thr Pro Ile Ala Gly
                85                  90                  95
```

```
Leu Asp Asn Ile Ile Leu Phe Leu Arg Gly Cys Lys Glu Leu Gly Leu
            100                 105                 110

Lys Glu Ser Gln Leu Phe Asp Pro Ser Asp Leu Gln Asp Thr Ser Asn
        115                 120                 125

Arg Val Thr Val Lys Ser Leu Asp Tyr Ser Arg Lys Leu Lys Asn Val
    130                 135                 140

Leu Val Thr Ile Tyr Trp Leu Gly Lys Ala Ala Asn Ser Cys Thr Ser
145                 150                 155                 160

Tyr Ser Gly Thr Thr Leu Asn Leu Lys Glu Phe Glu Gly Leu Leu Ala
                165                 170                 175

Gln Met Arg Lys Asp Thr Asp Asp Ile Glu Ser Pro Lys Arg Ser Ile
            180                 185                 190

Arg Asp Ser Gly Tyr Ile Asp Cys Trp Asp Ser Glu Arg Ser Asp Ser
        195                 200                 205

Leu Ser Pro Pro Arg His Gly Arg Asp Asp Ser Phe Asp Ser Leu Asp
    210                 215                 220

Ser Phe Gly Ser Arg Ser Arg Gln Thr Pro Ser Pro Asp Val Val Leu
225                 230                 235                 240

Arg Gly Ser Ser Asp Gly Arg Gly Ser Asp Ser Glu Ser Asp Leu Pro
                245                 250                 255

His Arg Lys Leu Pro Asp Val Lys Lys Asp Asp Met Ser Ala Arg Arg
            260                 265                 270

Thr Ser His Gly Glu Pro Lys Ser Ala Val Pro Phe Asn Gln Tyr Leu
        275                 280                 285

Pro Asn Lys Ser Asn Gln Thr Ala Tyr Val Pro Ala Pro Leu Arg Lys
    290                 295                 300

Lys Lys Ala Glu Arg Glu Glu Tyr Arg Lys Ser Trp Ser Thr Ala Thr
305                 310                 315                 320

Ser Pro Leu Gly Gly Glu Arg Pro Phe Arg Tyr Gly Pro Arg Thr Pro
                325                 330                 335

Val Ser Asp Asp Ala Glu Ser Thr Ser Met Phe Asp Met Arg Cys Glu
            340                 345                 350
```

```
Glu Glu Ala Ala Val Gln Pro His Ser Arg Ala Arg Gln Glu Gln Leu
        355                 360                 365

Gln Leu Ile Asn Asn Gln Leu Arg Glu Glu Asp Asp Lys Trp Gln Asp
    370                 375                 380

Asp Leu Ala Arg Trp Lys Ser Arg Arg Arg Ser Val Ser Gln Asp Leu
385                 390                 395                 400

Ile Lys Lys Glu Glu Glu Arg Lys Lys Met Glu Lys Leu Leu Ala Gly
                405                 410                 415

Glu Asp Gly Thr Ser Glu Arg Arg Lys Ser Ile Lys Thr Tyr Arg Glu
            420                 425                 430

Ile Val Gln Glu Lys Glu Arg Arg Glu Arg Glu Leu His Glu Ala Tyr
        435                 440                 445

Lys Asn Ala Arg Ser Gln Glu Glu Ala Glu Gly Ile Leu Gln Gln Tyr
    450                 455                 460

Ile Glu Arg Phe Thr Ile Ser Glu Ala Val Leu Glu Arg Leu Glu Met
465                 470                 475                 480

Pro Lys Ile Leu Glu Arg Ser His Ser Thr Glu Pro Asn Leu Ser Ser
                485                 490                 495

Phe Leu Asn Asp Pro Asn Pro Met Lys Tyr Leu Arg Gln Gln Ser Leu
            500                 505                 510

Pro Pro Pro Lys Phe Thr Ala Thr Val Glu Thr Thr Ile Ala Arg Ala
        515                 520                 525

Ser Val Leu Asp Thr Ser Met Ser Ala Gly Ser Gly Ser Pro Ser Lys
    530                 535                 540

Thr Val Thr Pro Lys Ala Val Pro Met Leu Thr Pro Lys Pro Tyr Ser
545                 550                 555                 560

Gln Pro Lys Asn Ser Gln Asp Val Leu Lys Thr Phe Lys Val Asp Gly
                565                 570                 575

Lys Val Ser Val Asn Gly Glu Thr Val His Arg Glu Glu Glu Lys Glu
            580                 585                 590

Arg Glu Cys Pro Thr Val Ala Pro Ala His Ser Leu Thr Lys Ser Gln
        595                 600                 605
```

```
Met Phe Glu Gly Val Ala Arg Val His Gly Ser Pro Leu Glu Leu Lys
    610                 615                 620

Gln Asp Asn Gly Ser Ile Glu Ile Asn Ile Lys Lys Pro Asn Ser Val
625                 630                 635                 640

Pro Gln Glu Leu Ala Ala Thr Thr Glu Lys Thr Glu Pro Asn Ser Gln
                645                 650                 655

Glu Asp Lys Asn Asp Gly Gly Lys Ser Arg Lys Gly Asn Ile Glu Leu
            660                 665                 670

Ala Ser Ser Glu Pro Gln His Phe Thr Thr Thr Val Thr Arg Cys Ser
        675                 680                 685

Pro Thr Val Ala Phe Val Glu Phe Pro Ser Ser Pro Gln Leu Lys Asn
    690                 695                 700

Asp Val Ser Glu Glu Lys Asp Gln Lys Lys Pro Glu Asn Glu Met Ser
705                 710                 715                 720

Gly Lys Val Glu Leu Val Leu Ser Gln Lys Val Val Lys Pro Lys Ser
                725                 730                 735

Pro Glu Pro Glu Ala Thr Leu Thr Phe Pro Phe Leu Asp Lys Met Pro
            740                 745                 750

Glu Ala Asn Gln Leu His Leu Pro Asn Leu Asn Ser Gln Val Asp Ser
        755                 760                 765

Pro Ser Ser Glu Lys Ser Pro Val Thr Thr Pro Gln Phe Lys Phe Trp
    770                 775                 780

Ala Trp Asp Pro Glu Glu Glu Arg Arg Arg Gln Glu Lys Trp Gln Gln
785                 790                 795                 800

Glu Gln Glu Arg Leu Leu Gln Glu Arg Tyr Gln Lys Glu Gln Asp Lys
                805                 810                 815

Leu Lys Glu Glu Trp Glu Lys Ala Gln Lys Glu Val Glu Glu Glu Glu
            820                 825                 830

Arg Arg Tyr Tyr Glu Glu Glu Arg Lys Ile Ile Glu Asp Thr Val Val
        835                 840                 845

Pro Phe Thr Val Ser Ser Ser Ser Ala Asp Gln Leu Ser Thr Ser Ser
    850                 855                 860
```

```
Ser Met Thr Glu Gly Ser Gly Thr Met Asn Lys Ile Asp Leu Gly Asn
865                 870                 875                 880

Cys Gln Asp Glu Lys Gln Asp Arg Arg Trp Lys Lys Ser Phe Gln Gly
                885                 890                 895

Asp Asp Ser Asp Leu Leu Leu Lys Thr Arg Glu Ser Asp Arg Leu Glu
            900                 905                 910

Glu Lys Gly Ser Leu Thr Glu Gly Ala Leu Ala His Ser Gly Asn Pro
        915                 920                 925

Val Ser Lys Gly Val His Glu Asp His Gln Leu Asp Thr Glu Ala Gly
    930                 935                 940

Ala Pro His Cys Gly Thr Asn Pro Gln Leu Ala Gln Asp Pro Ser Gln
945                 950                 955                 960

Asn Gln Gln Thr Ser Asn Pro Thr His Ser Ser Glu Asp Val Lys Pro
                965                 970                 975

Lys Thr Leu Pro Leu Asp Lys Ser Ile Asn His Gln Ile Glu Ser Pro
            980                 985                 990

Ser Glu Arg Arg Lys Lys Ser Pro  Arg Glu His Phe Gln  Ala Gly Pro
        995                 1000                1005

Phe Ser  Pro Cys Ser Pro Thr  Pro Pro Gly Gln Ser  Pro Asn Arg
    1010                1015                1020

Ser Ile  Ser Gly Lys Lys Leu  Cys Ser Ser Cys Gly  Leu Pro Leu
    1025                1030                1035

Gly Lys  Gly Ala Ala Met Ile  Ile Glu Thr Leu Asn  Leu Tyr Phe
    1040                1045                1050

His Ile  Gln Cys Phe Arg Cys  Gly Ile Cys Lys Gly  Gln Leu Gly
    1055                1060                1065

Asp Ala  Val Ser Gly Thr Asp  Val Arg Ile Arg Asn  Gly Leu Leu
    1070                1075                1080

Asn Cys  Asn Asp Cys Tyr Met  Arg Ser Arg Ser Ala  Gly Gln Pro
    1085                1090                1095

Thr Thr  Leu
    1100
```

<210> 46
<211> 1161
<212> PRT
<213> Homo sapiens

<400> 46

Met Lys Lys Phe Ser Arg Met Pro Lys Ser Glu Gly Gly Ser Gly Gly
1 5 10 15

Gly Ala Ala Gly Gly Gly Ala Gly Gly Ala Gly Ala Gly Ala Gly Cys
20 25 30

Gly Ser Gly Gly Ser Ser Val Gly Val Arg Val Phe Ala Val Gly Arg
35 40 45

His Gln Val Thr Leu Glu Glu Ser Leu Ala Glu Gly Gly Phe Ser Thr
50 55 60

Val Phe Leu Val Arg Thr His Gly Gly Ile Arg Cys Ala Leu Lys Arg
65 70 75 80

Met Tyr Val Asn Asn Met Pro Asp Leu Asn Val Cys Lys Arg Glu Ile
85 90 95

Thr Ile Met Lys Glu Leu Ser Gly His Lys Asn Ile Val Gly Tyr Leu
100 105 110

Asp Cys Ala Val Asn Ser Ile Ser Asp Asn Val Trp Glu Val Leu Ile
115 120 125

Leu Met Glu Tyr Cys Arg Ala Gly Gln Val Val Asn Gln Met Asn Lys
130 135 140

Lys Leu Gln Thr Gly Phe Thr Glu Pro Glu Val Leu Gln Ile Phe Cys
145 150 155 160

Asp Thr Cys Glu Ala Val Ala Arg Leu His Gln Cys Lys Thr Pro Ile
165 170 175

Ile His Arg Asp Leu Lys Val Glu Asn Ile Leu Leu Asn Asp Gly Gly
180 185 190

Asn Tyr Val Leu Cys Asp Phe Gly Ser Ala Thr Asn Lys Phe Leu Asn
195 200 205

Pro Gln Lys Asp Gly Val Asn Val Val Glu Glu Glu Ile Lys Lys Tyr
210 215 220

```
Thr Thr Leu Ser Tyr Arg Ala Pro Glu Met Ile Asn Leu Tyr Gly Gly
225                 230                 235                 240

Lys Pro Ile Thr Thr Lys Ala Asp Ile Trp Ala Leu Gly Cys Leu Leu
                245                 250                 255

Tyr Lys Leu Cys Phe Phe Thr Leu Pro Phe Gly Glu Ser Gln Val Ala
            260                 265                 270

Ile Cys Asp Gly Asn Phe Thr Ile Pro Asp Asn Ser Arg Tyr Ser Arg
        275                 280                 285

Asn Ile His Cys Leu Ile Arg Phe Met Leu Glu Pro Asp Pro Glu His
    290                 295                 300

Arg Pro Asp Ile Phe Gln Val Ser Tyr Phe Ala Phe Lys Phe Ala Lys
305                 310                 315                 320

Lys Asp Cys Pro Val Ser Asn Ile Asn Asn Ser Ser Ile Pro Ser Ala
                325                 330                 335

Leu Pro Glu Pro Met Thr Ala Ser Glu Ala Ala Ala Arg Lys Ser Gln
            340                 345                 350

Ile Lys Ala Arg Ile Thr Asp Thr Ile Gly Pro Thr Glu Thr Ser Ile
        355                 360                 365

Ala Pro Arg Gln Arg Pro Lys Ala Asn Ser Ala Thr Thr Ala Thr Pro
    370                 375                 380

Ser Val Leu Thr Ile Gln Ser Ser Ala Thr Pro Val Lys Val Leu Ala
385                 390                 395                 400

Pro Gly Glu Phe Gly Asn His Arg Pro Lys Gly Ala Leu Arg Pro Gly
                405                 410                 415

Asn Gly Pro Glu Ile Leu Leu Gly Gln Gly Pro Pro Gln Gln Pro Pro
            420                 425                 430

Gln Gln His Arg Val Leu Gln Gln Leu Gln Gln Gly Asp Trp Arg Leu
        435                 440                 445

Gln Gln Leu His Leu Gln His Arg His Pro His Gln Gln Gln Gln Gln
    450                 455                 460

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
465                 470                 475                 480
```

```
Gln Gln Gln Gln Gln Gln His His His His His His His His Leu Leu
                485                 490                 495

Gln Asp Ala Tyr Met Gln Gln Tyr Gln His Ala Thr Gln Gln Gln Gln
            500                 505                 510

Met Leu Gln Gln Gln Phe Leu Met His Ser Val Tyr Gln Pro Gln Pro
        515                 520                 525

Ser Ala Ser Gln Tyr Pro Thr Met Met Pro Gln Tyr Gln Gln Ala Phe
    530                 535                 540

Phe Gln Gln Gln Met Leu Ala Gln His Gln Pro Ser Gln Gln Gln Ala
545                 550                 555                 560

Ser Pro Glu Tyr Leu Thr Ser Pro Gln Glu Phe Ser Pro Ala Leu Val
                565                 570                 575

Ser Tyr Thr Ser Ser Leu Pro Ala Gln Val Gly Thr Ile Met Asp Ser
            580                 585                 590

Ser Tyr Ser Ala Asn Arg Ser Val Ala Asp Lys Glu Ala Ile Ala Asn
        595                 600                 605

Phe Thr Asn Gln Lys Asn Ile Ser Asn Pro Pro Asp Met Ser Gly Trp
    610                 615                 620

Asn Pro Phe Gly Glu Asp Asn Phe Ser Lys Leu Thr Glu Glu Glu Leu
625                 630                 635                 640

Leu Asp Arg Glu Phe Asp Leu Leu Arg Ser Asn Arg Leu Glu Glu Arg
                645                 650                 655

Ala Ser Ser Asp Lys Asn Val Asp Ser Leu Ser Ala Pro His Asn His
            660                 665                 670

Pro Pro Glu Asp Pro Phe Gly Ser Val Pro Phe Ile Ser His Ser Gly
        675                 680                 685

Ser Pro Glu Lys Lys Ala Glu His Ser Ser Ile Asn Gln Glu Asn Gly
    690                 695                 700

Thr Ala Asn Pro Ile Lys Asn Gly Lys Thr Ser Pro Ala Ser Lys Asp
705                 710                 715                 720

Gln Arg Thr Gly Lys Lys Thr Ser Val Gln Gly Gln Val Gln Lys Gly
                725                 730                 735
```

```
Asn Asp Glu Ser Glu Ser Asp Phe Glu Ser Asp Pro Pro Ser Pro Lys
            740                 745                 750

Ser Ser Glu Glu Glu Glu Gln Asp Asp Glu Glu Val Leu Gln Gly Glu
        755                 760                 765

Gln Gly Asp Phe Asn Asp Asp Asp Thr Glu Pro Glu Asn Leu Gly His
    770                 775                 780

Arg Pro Leu Leu Met Asp Ser Glu Asp Glu Glu Glu Glu Lys His
785                 790                 795                 800

Ser Ser Asp Ser Asp Tyr Glu Gln Ala Lys Ala Lys Tyr Ser Asp Met
                805                 810                 815

Ser Ser Val Tyr Arg Asp Arg Ser Gly Ser Gly Pro Thr Gln Asp Leu
            820                 825                 830

Asn Thr Ile Leu Leu Thr Ser Ala Gln Leu Ser Ser Asp Val Ala Val
        835                 840                 845

Glu Thr Pro Lys Gln Glu Phe Asp Val Phe Gly Ala Val Pro Phe Phe
    850                 855                 860

Ala Val Arg Ala Gln Gln Pro Gln Gln Glu Lys Asn Glu Lys Asn Leu
865                 870                 875                 880

Pro Gln His Arg Phe Pro Ala Ala Gly Leu Glu Gln Glu Glu Phe Asp
                885                 890                 895

Val Phe Thr Lys Ala Pro Phe Ser Lys Lys Val Asn Val Gln Glu Cys
            900                 905                 910

His Ala Val Gly Pro Glu Ala His Thr Ile Pro Gly Tyr Pro Lys Ser
        915                 920                 925

Val Asp Val Phe Gly Ser Thr Pro Phe Gln Pro Phe Leu Thr Ser Thr
    930                 935                 940

Ser Lys Ser Glu Ser Asn Glu Asp Leu Phe Gly Leu Val Pro Phe Asp
945                 950                 955                 960

Glu Ile Thr Gly Ser Gln Gln Gln Lys Val Lys Gln Arg Ser Leu Gln
                965                 970                 975

Lys Leu Ser Ser Arg Gln Arg Arg Thr Lys Gln Asp Met Ser Lys Ser
            980                 985                 990
```

```
Asn Gly Lys Arg His His Gly Thr  Pro Thr Ser Thr Lys  Lys Thr Leu
        995                 1000                 1005

Lys Pro  Thr Tyr Arg Thr Pro  Glu Arg Ala Arg Arg  His Lys Lys
    1010                1015                1020

Val Gly  Arg Arg Asp Ser Gln  Ser Ser Asn Glu Phe  Leu Thr Ile
    1025                1030                1035

Ser Asp  Ser Lys Glu Asn Ile  Ser Val Ala Leu Thr  Asp Gly Lys
    1040                1045                1050

Asp Arg  Gly Asn Val Leu Gln  Pro Glu Glu Ser Leu  Leu Asp Pro
    1055                1060                1065

Phe Gly  Ala Lys Pro Phe His  Ser Pro Asp Leu Ser  Trp His Pro
    1070                1075                1080

Pro His  Gln Gly Leu Ser Asp  Ile Arg Ala Asp His  Asn Thr Val
    1085                1090                1095

Leu Pro  Gly Arg Pro Arg Gln  Asn Ser Leu His Gly  Ser Phe His
    1100                1105                1110

Ser Ala  Asp Val Leu Lys Met  Asp Asp Phe Gly Ala  Val Pro Phe
    1115                1120                1125

Thr Glu  Leu Val Val Gln Ser  Ile Thr Pro His Gln  Ser Gln Gln
    1130                1135                1140

Ser Gln  Pro Val Glu Leu Asp  Pro Phe Gly Ala Ala  Pro Phe Pro
    1145                1150                1155

Ser Lys  Gln
    1160
```